Published in final edited form as:

Sociol Methodol. 2014 August ; 44(1): 229–272. doi:10.1177/0081175013490188.

Comparing Spatial and Multilevel Regression Models for Binary Outcomes in Neighborhood Studies

Hongwei Xu

Survey Research Center, University of Michigan, 426 Thompson St, 216 NU ISR Bldg, Ann Arbor, MI 48104, Telephone: +1 (734) 615-3552, Fax: +1 (734) 763-1428, xuhongw@umich.edu

Abstract

The standard multilevel regressions that are widely used in neighborhood research typically ignore potential between-neighborhood correlations due to underlying spatial processes, and hence produce inappropriate inferences about neighborhood effects. In contrast, spatial models make estimations and predictions across areas by explicitly modeling the spatial correlations among observations in different locations. A better understanding of the strengths and limitations of spatial models as compared to the standard multilevel model is needed to improve the research on neighborhood and spatial effects. This research systematically compares model estimations and predictions for binary outcomes between (distance- and lattice-based) spatial and the standard multilevel models in the presence of both within- and between-neighborhood correlations, through simulations. Results from simulation analysis reveal that the standard multilevel and spatial models produce similar estimates of fixed effects, but different estimates of random effects variances. Both the standard multilevel and pure spatial models tend to overestimate the corresponding random effects variances, compared to hybrid models when both non-spatial within neighborhood and spatial between-neighborhood effects exist. Spatial models also outperform the standard multilevel model by a narrow margin in case of fully out-of-sample predictions. Distance-based spatial models provide extra spatial information and have stronger predictive power than lattice-based models under certain circumstances. These merits of spatial modeling are exhibited in an empirical analysis of the child mortality data from 1880 Newark, New Jersey.

Keywords

Spatial statistics; multilevel regression; Bayesian inference; neighborhood effects; binary outcome

1. INTRODUCTION

Multilevel regression analysis 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 standard multilevel model helps correct for within-neighborhood correlation among individual observations and thus adjusts for standard errors, resulting in efficient estimates for individual- and neighborhood-level predictors (Diggle et al. 2002). It also allows an assessment of within- and between-neighborhood variations (Snijders and

Bosker 1994) as well as how individual- and neighborhood-level predictors contribute to these variations (Diez-Roux 2000). Nevertheless, the standard multilevel model typically ignore potential between-neighborhood correlations due to spatial diffusion processes for example, and assume independent observations in one neighborhood from those in another neighborhood, which may lead to the overstatement of the statistical significance of neighborhood effects (Chaix, Merlo, and Chauvin 2005).

The left-hand-side diagram in Figure 1 illustrates a hypothetical example of the assumption of within-neighborhood correlation within the standard multilevel model. Each cell in the grid represents a neighborhood and its color indicates the average level (from low to high as indicated by darker colors) of a certain individual outcome shared by the observations from that neighborhood. Within a neighborhood, each observation's outcome deviates around the neighborhood's mean. Neighborhood mean levels of a given 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 the left-hand-side diagram in Figure 1). The seemingly random distribution of mean outcomes at the neighborhood-level across the entire study area reflects the assumption of between-neighborhood independence. That is, the mean outcome is no more similar between two adjacent neighborhoods than between two neighborhoods far away from each other.

In reality, however, between-neighborhood correlations may exist as a function of the distance between two nearby neighborhoods as stated in Tobler's First Law of Geography, "Everything is related to everything else, but near things are more related than distant things." (1970: 236) First of all, a neighborhood's socioeconomic and political resources are likely to be linked to those in adjacent neighborhoods within a larger 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 et al. 2008; Guo and Bhat 2007; Riva et al. 2008; Tatalovich et al. 2006). Instead, they may transcend neighborhood boundaries and thus be affected by or consequential to what happens in nearby areas (Liu, Wall, and Hodges 2005; Sampson et al. 2002). For example, collective efficacy in a neighborhood has been found to benefit residents living in adjacent neighborhoods (Sampson et al. 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).

Thirdly, a spatial diffusion process may operate either within a city-wide system or even on a much larger geographic scale such that information, knowledge, techniques, innovations, products, languages, and diseases likely to spread from one place to another. For instance, drawing on locational information for the incidence of homicide and participants' residences, Cohen and Tita (1999) found a spatial spread of homicides from gang youth in one census tract to non-gang youth across neighboring tracts in the city of Pittsburgh during the period 1991 to 1995. On an even a greater geographic scale, Hedström (1994) examined the growth of 16,911 trade unions that were distributed over districts across Sweden during

the period 1890 to 1940. He found a significant association of union formations in a district with union activities, weighted by distance, in other districts. In short, exposure to social and environmental circumstances in one neighborhood can be correlated with those in another.

Spatial models have been developed to make estimations and predictions about space by explicitly modeling the spatial correlations among observations in different locations (Diggle et al. 1998). Two of the key goals of spatial analysis are 1) to estimate the spatial distribution of an outcome of interest across the study area based on observations at a discrete set of locations and 2) to make predictions at a new location (Diggle et al. 2003). These model estimations and predictions 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 outcomes of interest while adjusting for non-independent observations (Chaix et al. 2005; Dietz 2002).

The right-hand side of the diagram in Figure 1 illustrates a hypothetical example of the assumption of the 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 that the strength of correlation between two locations declines as the distance between them increases, resulting in similar mean levels of outcomes among nearby neighborhoods and hence clusters of neighborhoods with similar mean outcomes (as indicated by the color gradient across the cells in the right-hand side of the diagram in Figure 1).

Alternatively, spatial correlations across neighborhoods can be incorporated into a standard multilevel model in a variety of different ways, including multiple membership relationships in which individuals are involved (Browne, Goldstein, and Rasbash 2001), a conditional autoregressive structure among nearby neighborhoods with extra place effects (Arcaya, Brewster, Zigler, and Subramanian 2012), and other forms of correlated spatial random effects (Browne and Goldstein 2010). In essence, these models take a similar approach as that proposed by Diggle and colleagues (1998), that is, extending the standard multilevel model by allowing extra spatial correlations across the neighborhood boundaries.

Chaix and colleagues (2005) are among the first to compare the spatial approach with the standard multilevel approach for studying neighborhood effects on health. Through an empirical analyses of healthcare utilization in France, Chaix and colleagues (2005) demonstrated that the standard multilevel model can 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 whether they can be generalized to other research settings. Moreover, they did not provide a thorough comparison of model performance in terms of both model estimation and prediction between spatial and standard multilevel models through a formal approach such as simulation analysis (Burton et al. 2006).

While researchers have become increasingly interested in the spatial dynamics beyond simple neighborhood-level variation (Dietz 2002; Logan et al. 2010; Sampson et al. 2002), the potential utility of spatial models in neighborhood studies remains underappreciated despite the recent advancement in spatial techniques and increased availability of spatial data. Thus, a better understanding of the strengths and weaknesses of spatial models as compared to the standard multilevel model is urgently needed to provide methodological guidance for empirical studies to avoid erroneous statistical inferences and substantive conclusions. The main purpose of this paper is to assess the impact of disregarding extant spatial correlations across neighborhoods when studying neighborhood effects on individual outcomes. The present study focuses on models of binary outcome using a logit link because of their increased prevalence and popularity in neighborhood and spatial studies. By systematically comparing model performance in estimation and prediction between spatial and the standard multilevel models in the presence of both within- and between-neighborhood correlations through simulation analyses, this study informs researchers about making cautious model choices whenever spatial information is available, in addition to standard multilevel data structures. Drawing upon a real life data set, this study also illustrates the application of the spatial approach in empirical research and further demonstrates its relative advantages compared to the standard multilevel approach.

2. MULTILEVEL AND SPATIAL MODELING

2.1. Standard Multilevel Model

Let y_{ij} denote the binary outcome for an individual i in neighborhood j , and assume y_{ij} follows a Bernoulli distribution with success probability p_{ij} or Binomial(1, p_{ij}). Using an appropriate link function such as logit, a binary outcome can be associated with linear predictors as the following,

$$\text{logit}[E(y_{ij})] = \text{logit}(p_{ij}) = \alpha_0 + X_{ij}\beta + Z_j\gamma + u_j \quad (1)$$

where α_0 is the regular intercept, $X_{ij}\beta$ is the product of individual-level predictors and the corresponding unknown parameters, and $Z_j\gamma$ is the product of neighborhood-level predictors and the associated 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 (for details on standard multilevel logit models, see Goldstein 2010).

2.2. Pure Spatial Model

By contrast, a pure spatial model ignores within-neighborhood but incorporates between-neighborhood correlations in the following way,

$$\text{logit}(p_{ij}) = \alpha_0 + X_{ij}\beta + Z_j\gamma + s_j \quad (2)$$

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

$$s \sim N(0, \sigma_s^2 H(\phi)) \quad (3)$$

where σ_s^2 denotes the variance of the spatial random effects, known as the *partial sill* in the spatial literature (Banerjee et al. 2004). The other component, denoted by $H(\phi)$, 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 et al. 2005) or when a large number of different locations are computationally too expensive to be fully incorporated (Gelfand et al. 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,

$$H(\phi)_{ij} = \rho(d_{ij}, \phi) \quad (4)$$

where ρ is typically chosen to be an isotropic function, which assumes that the correlation between two locations only depends 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 the two locations increases. The distance at which the spatial correlation drops to 5 percent and can be considered as “no longer existing” is known as the *effective range* in the literature (Banerjee et al. 2004).

A common choice for the isotropic spatial correlation is the following exponential function because of its relatively simple form and hence relatively low computational cost as well as its wide availability in statistical packages,

$$H(\phi)_{ij} = \exp(-\phi d_{ij}) \quad (5)$$

Setting $\exp(-\phi d_{ij})$ equal to 0.05 and solving the equation for d_{ij} , it is straightforward to see that in this case, the effective range is approximately $3/\phi$. Several other types of spatial correlation functions have also been proposed in the literature (see Chapter 2 in Banerjee et al. 2004). For example, a Gaussian function results from raising the power of the product ϕd_{ij} to 2, that is,

$$H(\phi)_{ij} = \exp[-(\phi d_{ij})^2] \quad (6)$$

Setting $\exp[-(\phi d_{ij})^2]$ equal to 0.05 and solving the equation for d_{ij} , it is straightforward to see that in this case, the effective range is approximately $\sqrt{3}/\phi$. However, a serious problem of model convergence was encountered when using a Gaussian function in the exploratory simulation analysis as there is often very little information in the data to estimate ϕ (Thomas et al. 2004). Therefore, this study mainly focuses on the exponential function and discusses supplementary results from using an approximately linear correlation function which has a less severe problem of model convergence and takes the following form,

$$H(\alpha)_{ij} = \frac{2}{\pi} \left\{ \cos^{-1} \left(\frac{d_{ij}}{\alpha} \right) - \left[\frac{d_{ij}}{\alpha} \left(1 - \frac{d_{ij}^2}{\alpha^2} \right) \right]^{\frac{1}{2}} \right\} \text{ for } d_{ij} < \alpha \quad (7)$$

and $H(\alpha)_{ij} = 0$ for $d_{ij} > \alpha$. A large value of α dictates a slow rate of an approximately linear decline of correlation with increasing distance between two units, with correlation dropping to zero at a distance equal to α .

2.3. Hybrid Model

Combining the standard multilevel model and the pure spatial model above, Diggle and colleagues (1998) are among the first to develop a generalized linear spatial model as the following,

$$\text{logit}(p_{ij}) = \alpha_0 + X_{ij}\beta + Z_j\gamma + u_j + s_j \quad (8)$$

where u_j and s_j capture within- and between-neighborhood correlations as specified in Equations (1) and (2), respectively. Details on parameter estimation and prediction using either maximum likelihood estimation or Bayesian inference can be founded in the work by Diggle and colleagues (1998; 2003).

Since then, other scholars have proposed alternative models that can simultaneously accommodate within- and between-neighborhood correlations in different ways. For example, Browne and Goldstein (2010: 454) developed a modeling framework with correlated random effects that takes into account the possibility that “some pairs of clusters are more similar to each other than to other clusters” and meanwhile allows within-cluster correlations. Their modeling strategy takes essentially the same form as specified in Equation (8), that is, a linear combination of a non-spatially structured correlation and a spatially structured correlation (i.e. $u_j + s_j$). In fact, dropping s_j from the right hand side of Equation (1) results in the standard multilevel model known as the 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.

Browne and Goldstein (2010) distinguish two ways to capture s_j , the spatial correlation between clusters or neighborhoods in the context of the present study. The first one explicitly models the correlation as a function of distance between neighborhoods, which follows the same approach as specified in Equations (3) and (4), although Browne and Goldstein (2010) chose functions such as inverse hyperbolic, logit, and log that are less common in the spatial analysis literature.

The second way to capture s_j is relatively less spatially informed, and involves creating a lattice-like structure linking adjacent neighborhoods. A widely used specification leads to the intrinsic conditional autoregressive (CAR) model (Besag et al. 1991) with the conditional distribution of s_i defined as,

$$s_i | s_{-i} \sim N\left(\sum_{j \in A(i)} \frac{s_j}{n_{A(i)}}, \frac{\sigma_s^2}{n_{A(i)}}\right) \quad (9)$$

where s_{-i} denotes the rest of the study area excluding the i th neighborhood, $A(i)$ represents the adjacent area of the i th neighborhood, $n_{A(i)}$ is the number of adjacent neighborhoods, and

hence the mean component is an average over the adjacent area of the i th neighborhood. This is also the approach demonstrated and recommended by Arcaya and colleagues (2012).

In their original example of county-level life expectancy, a continuous outcome, Arcaya and colleagues (2012) formulated a multilevel-CAR model that simultaneously accounts for both place effects at the state-level and space effects induced by each county's adjacent counties ignoring state boundaries. With minor modifications of the hierarchical structure, their model can be adapted to the one as specified in Equation (8), where individuals are nested within neighborhoods (inducing place effects), neighborhood-level random effects are captured by u_j , and between-neighborhood correlations (inducing space effects) are accounted for by a CAR specification for each neighborhood's adjacent neighborhood units as in Equation (9).

The CAR model is closely related to the multiple-membership model described by Browne et al. (2001). Modifying their original example of count of male lip cancer in 56 regions of Scotland by considering a binary outcome for individuals nested within neighborhoods instead, their model can be written as follows,

$$\text{logit}(p_{i,j}) = \alpha_0 + X_{i,j}\beta + u_j^{(2)} + \sum_{k \in A(j)} w_{j,k}^{(3)} u_k^{(3)} \quad (10.1)$$

$$u_j^{(2)} \sim N(0, \sigma_{u(2)}^2), \quad u_k^{(3)} \sim N(0, \sigma_{u(3)}^2), \quad w_{j,k}^{(3)} = \frac{1}{n_{A(i)}} \quad (10.2)$$

where $p_{i,j}$ denotes the outcome for individual i living in neighborhood j ; membership classification 1 ("identity" classification) applies to the lowest level – individuals; $u_j^{(2)}$ represents the effects of membership classification 2 – the neighborhood where the i th individual resides; the k th adjacent neighborhood for individual i , $u_k^{(3)}$ represents the effects of membership classification 3 – the k th adjacent neighborhood of the focal neighborhood j where individual i resides; and $w_{j,k}^{(3)}$ is a corresponding weight. The main difference between the multiple-membership model and the CAR model lies in that the latter incorporates spatial correlation through a variance structure instead of a multiple membership relationship where the neighborhood random effects are assumed independent (Browne, Goldstein, and Rasbash 2001).

The present study focuses on comparing four sets of models, including the *standard multilevel model* (as the reference model), the *pure spatial model* with distance-based correlations between neighborhoods, and two *hybrid models* with one using lattice-like CAR structure and the other using more spatially explicit distance-based correlations as in the pure spatial model. Even though maximum likelihood-based estimates have been developed for both multilevel and spatial models, Bayesian inference is adopted in the present study for its relative merits. First, with the booming advancement in Markov Chain Monte Carlo (MCMC) (Gilks et al. 1996) and data augmentation (Tanner and Wong 1987) algorithms in the recent decades, Bayesian analyses are highly flexible in expanding a standard multilevel model to accommodate more complex hierarchical (e.g. spatial) data structures without

spending extra efforts in modifying existing estimation procedures. As a result, Bayesian inference has been widely applied in spatial statistics (e.g. Banerjee et al. 2004) and hence I made the choice in the present study to be consistent with the existing literature (Arcaya et al. 2012; Browne and Goldstein 2010; Browne et al. 2001; Chaix et al. 2005; Henderson et al. 2002). Second, the extension from coefficient estimation to out-of-sample prediction can also be easily accomplished in a Bayesian framework by adding another hierarchical level, that is, simulating predictive distributions that are conditional upon estimated posterior distributions of parameters. Thus, Bayesian predictions naturally incorporate the uncertainty inherent in parameter estimation from a sample of data which often requires extra work to be appropriately addressed when using a maximum likelihood approach (Diggle and Ribeiro 2002).

3. SIMULATION ANALYSIS

The simulation analysis here compares spatial and the standard multilevel models with respect to parameter estimation and prediction in the presence of both within- and between-neighborhood correlations. As the true data generation process and values of parameters are known *a priori*, which is never the case in any empirical study, the simulation analysis enables a direct assessment of the appropriateness and accuracy of a variety of models (Burton et al. 2006). To reduce the computational burden, the simulation analysis in this study focuses on the case of only one independent variable at the individual and one independent variable at the neighborhood level. By varying the relative strengths of the within- and between-neighborhood correlations, the simulation study demonstrates the advantages and weaknesses of different spatial models in relation to the standard multilevel model that ignores spatial correlations, and hence provides guidance for model selection in empirical studies.

The gold standard K -fold cross-validation (Kohavi 1995) was employed in the simulation analysis. Simulation analyses which involve fitting the training data alone are susceptible to the issue of overfitting. The same model very often reaches a relatively poor goodness-of-fit for an independent sample of the validation data from the same population as the training data. By partitioning a simulated data set into K subsets and rotationally leaving one subset out as validation data, the K -fold cross validation allows a simple yet effective assessment of the predictive power across models, that is, to what extent the results of a model are generalizable to an independent data set, in addition to evaluating parameter estimation. 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 difficulties associated with using other usual methods such as residual plots (Gelman et al. 2000). For each cross-validation data set, predictions are made for the validation observations with “missing” binary outcomes based on the model fitted to the training data. A comparison of the predicted values with the true values provides an easy yet straightforward way of assessing the accuracy of a predictive model in practice (i.e. out-of-sample prediction) while adjusting for uncertainty in estimating the model parameters.

Given that the assumption of distance-decay correlation underlies most spatial analyses and that the distance-based spatial correlation model remains something of an underdog in

empirical research, it serves the “true” data generation mechanism in the present study. Taking exponential spatial correlation structure as an example, 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 a mean 0, a partial sill σ_s^2 , and a decay parameter ϕ . The neighborhood structure is represented by an 8×8 grid with 64 neighborhoods in total (see Figure 2).
2. Neighborhood-level random effects are randomly generated from a normal distribution with a mean 0 and a variance σ_u^2 (i.e. nugget).
3. A neighborhood-level predictor (i.e. neighborhood-level fixed effects), z , is sampled from a standard normal distribution across 64 neighborhoods. Within each neighborhood, an individual-level predictor (i.e. individual-level fixed effects), x , 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 for each observation.
4. For each observation, the values of fixed effects, neighborhood random effects, 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 distribution (for logit model) or a normal distribution (for probit model) depending on which link function to be used.
5. A 4-fold cross validation is constructed by randomly partitioning a simulated data set as described above into four mutually exclusive subsamples, each of which contains 480 observations. Each of the four subsamples is used in turn as the so-called validation data. That is, the outcome values of its 480 observations will be treated as unknown, while a model will be fitted to the other three subsamples with 1440 observations, known as the training data. Predictions for the 480 observations in the validation data will be made based on the fitted models and compared to their true values to assess the model performance.

To summarize, the mean response for the i th observation in the j th neighborhood was generated according to the following model with both within- and between-neighborhood correlations:

$$g(\mu_{ij}) = \beta_0 + x_{ij}\beta_1 + Z_j\gamma_1 + u_j + s_j \quad (11.1)$$

$$u_j \sim N(0, \sigma_u^2) \quad (11.2)$$

$$s_j \sim N(0, \sigma_s^2 \exp(-\phi d_{ij})) \quad (11.3)$$

Throughout the simulations, the true values for the fixed effects parameters are kept as $\beta_0 = -0.5$, $\beta_1 = 0.8$, and $\gamma_1 = -0.5$, whereas those for the random effects parameters are varied to

reflect four scenarios that differ in the relative strengths of spatial correlation to within-neighborhood correlation. In Scenario 1, using exponential function, for example, for spatial correlation and replacing the parameters in the equations above with a set of true values, the data are simulated according to:

$$g(\mu_{ij}) = -0.5 + 0.8 * x_{ij} - 0.5 * z_j + u_j + s_j \quad (12.1)$$

$$x_{ij} \sim N(0, 1) \quad (12.2)$$

$$z_j \sim N(0, 1) \quad (12.3)$$

$$u_j \sim N(0, 3) \quad (12.4)$$

$$s_j \sim (0, 3 * \exp(-3d_{ij})) \quad (12.5)$$

In order to mimic real data more closely, extra noise was added to the linear predictors in Equation (12.1) by making a random draw from $N(0, 1)$. The seemingly large values of σ_u^2 and σ_s^2 were taken to offset this added extra random variation and to be comparable to the equally large coefficients of β_1 and γ_1 . In addition, given that only two predictors are considered in the simulation, which is unlikely to represent the complex reality that often involves more variables, the relatively large values of σ_u^2 and σ_s^2 can be considered to absorb the unexplained variation contributed by omitted variables in the observed data. The decay parameter for the spatial correlation is set to $\phi = 3$ such that the effective range is 1, the side length of the simulated rectangular study area (see Figure 2). Thus, Scenario 1 ($\sigma_u^2=3, \sigma_s^2=3$, and $\phi = 3$) mimics a situation where between-neighborhood correlation is as strong at distance 0 as within-neighborhood correlation, and it disappears relatively slowly as the distance between two neighborhoods increases.

Scenario 2 applies the same values of $\sigma_u^2(=3)$ and $\phi (= 3)$, but reduces σ_s^2 by half to be 1.5, simulating a situation where between-neighborhood correlation is only half as strong at distance 0 as within-neighborhood correlation, but it remains weakly effective at a long range. Scenario 3 mimics a situation with relatively strong spatial correlation which, however, has only half as long as the effective range in Scenario 1 and 2 by using $\sigma_u^2=3, \sigma_s^2=3$, and $\phi = 6$. Scenario 4 applies $\sigma_u^2=3, \sigma_s^2=1.5$, and $\phi = 6$, and hence simulating a situation with weak spatial correlation and short effect range. These four scenarios together permit a more thorough assessment of the relative strengths and weaknesses of the four different models. It is worth noting that when the data are simulated based upon Gaussian spatial correlation function, the true values of ϕ are set as $\sqrt{3}$ and $\sqrt{6}$ to mimic long and short effective ranges, respectively, as they correspond to the same distances (1 and 0.5) at which Gaussian spatial correlation almost disappears (see Equation (6)) as when using exponential function. It is also worth mentioning that even under Scenario 1 (strong spatial correlation and long effective range), the correlation between two adjacent neighborhoods is

still set to be weaker than that within the same neighborhood, a more appropriate approximation of the reality.¹ Therefore, a pure spatial model that incorporates spatial correlation but overlooks within-neighborhood correlation may be superior to a standard multilevel model, but should not be expected to outperform a hybrid model.

The validation data set is constructed in two different ways. First, as is conventionally done, the 480 observations are randomly selected across the entire study area so that in each subsample of the 4-fold cross validation, each neighborhood is likely to have some observations in the validation data and in the training data. This approach allows model comparisons in terms of making predictions for “new” observations with different individual attributes but the same neighborhood attributes and spatial location information as those in the training data. Thereby, the first approach is hereinafter referred to as “partially out-of-sample” cross validation.

Second, instead of randomly choosing individual observations, 16 neighborhoods are randomly selected for each subsample of the 4-fold cross validation. All the observations (480 in total) in these selected neighborhoods are assigned missing outcome values and treated as validation data. In contrast to the first approach, the second approach is focused on comparing the predictive power of “new” observations with not only individual but also neighborhood attributes and spatial locations that are different from those in the training data. Thus, the second approach is referred to as “fully out-of-sample” cross validation.

For both partially and fully out-of-sample cross validation, a total number of 100 data sets were simulated for each set of parameter values. Each data set was partitioned into 4 subsamples, resulting in 400 data sets in total for model fitting. Four models were fitted to each simulated data set and compared, including a *standard multilevel model* that adjusts for within-neighborhood correlation, a *pure spatial model* that adjusts for distance-based between-neighborhood correlation but ignores within-neighborhood correlation, and two hybrid models that adjust for both within- and between-neighborhood correlations. The first hybrid model accommodates the spatial correlation through lattice-based CAR distribution as specified in Equation (9), and is referred to as the *hybrid CAR* model. The second one models distance-based between-neighborhood correlation, and is thus the “true” model in the simulation analysis, and is referred to as the *hybrid spatial* model. The standard multilevel model mainly serves as a benchmark for assessing the other three models.

Model estimation and prediction were carried out by using OpenBUGS version 3.2.2 (Lunn et al. 2009), an open-source software package for performing Bayesian inference using MCMC algorithms. To ensure model convergence, each model was fitted by initiating 2 MCMC chains with different starting values and letting each chain run for 60,000 iterations. Model convergence was monitored by graphically examining the trace plots of MCMC

¹The simulated study area is an 8×8 grid with each cell of $1/7$ (≈ 0.143) unit in side length since the distance between the centroid of the leftmost cell and that of the rightmost cell in the same row is of 1 unit (see Figure 2). Thus, the shortest distance between any two neighborhoods is $1/7$. Under Scenario 1, where both the variance of the spatial random effects (σ_s^2) and the distance decay parameter (ϕ) are set to be 3, the variance for the spatial correlation (s_j) between observations from the two nearest neighborhoods is only about $1.95 (= \sigma_s^2 \exp(-\phi d) = 3 * \exp(-3 * 1/7))$, only about two thirds of the variance for the within-neighborhood correlation (σ_u^2) which is set to be 3.

chains as well as computing the Gelman-Rubin statistic (Brooks and Gelman 1998; Gelman and Rubin 1992), a measure for assessing the convergence of multiple chains based on between- and within-chain variances. After discarding the first 30,000 iterations as the burn-in, each chain was then thinned by storing the sampled parameter values from every fiftieth iteration in order to reduce its autocorrelation, which resulted in a total number of 1,000 iterations from the 2 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 parameter estimation from different models (Burton, Altman, Royston, and Holder 2006). Bias is assessed by the difference, calculated as a percentage of the true value, between the average estimate and the true value, known as the percentage bias (PB). Coverage is assessed by the proportion of times the true parameter value falls within the estimated 95% credible interval, known as the coverage rate (CR). Accuracy is assessed by the mean-square error (MSE) which is a combined measure of bias and variability. The relative goodness-of-fit of a model was measured by the deviance information criterion (DIC) (Spiegelhalter et al. 2002), a hierarchical modeling generalization of the Akaike information criterion (AIC) widely used to compare non-nested regression models (Akaike 1974). A smaller value of the DIC indicates a better model fit to the data.

4. SIMULATION RESULTS

This section first assesses model performance in terms of parameter estimation and prediction based on partially and fully out-of-sample cross validation, and then evaluates robustness against model misspecifications from fully out-of-sample cross validation.

4.1. Partially versus Fully Out-of-Sample Cross Validation

Table 1 presents measures of parameter estimation from logit models fitted to data generated from a process in which spatial correlation follows an exponential distance-decay function. For partially out-of-sample analysis, only the results from Scenario 1 (i.e. strong spatial correlation and long effective range) are shown as those from other scenarios are largely similar. Under Scenario 1, the four models perform somewhat similarly with respect to the accuracy, precision, and efficiency of the estimation of the fixed effects (i.e., β_0 , β_1 , and γ_1), regardless of the data being generated from partially or fully out-of-sample cross validation. One notable finding (of relatively little substantive interest) is that compared to the standard multilevel and hybrid CAR models, the pure and hybrid spatial models produce less accurate estimates, as measured by MSE and PB, of β_0 while simultaneously being more likely to cover its true value with a 95 percent credible interval, that is, an interval with the 2.5 and 97.5 percentiles of the posterior distribution as its lower and upper limits, respectively. Moreover, in the case of fully out-of-sample cross validation, the coefficient estimate for the neighborhood-level predictor (γ_1) is relatively less biased in the standard multilevel and

hybrid CAR models than that in the pure and hybrid spatial models, although the four models do not differ substantially as measured by CR and MSE.

Remarkable variations appear across the four different types of models in their estimation of the random effects variances. With respect to σ_w^2 , the variance of within-neighborhood correlation, the two hybrid models outperform the standard multilevel model. Specifically, in the case of partially out-of-sample cross validation, the hybrid CAR model was slightly less biased than the hybrid spatial model, although they both perform better in terms of accuracy and precision compared to the standard multilevel model. In the case of fully out-of-sample cross validation, the hybrid spatial model provides the most accurate estimate as indicated by its smallest MSE, although the hybrid CAR model leads to the least biased estimate. These three models do not differ considerably from each other in coverage rate, although the two hybrid models maintain some minor superiority in this regard. Turning to σ_s^2 , the variance of between-neighborhood correlation, the hybrid spatial model provides the least biased estimate with the highest CR as well as much better precision compared to the hybrid CAR model which adopts a lattice structure and hence disregards extra spatial information. The estimate from pure spatial model has marginally smaller MSE than that from the hybrid spatial model, but it performs quite poorly in terms of CR compared to the two hybrid models. The hybrid spatial model outperforms the pure spatial one with less biased and more accurate estimates of ϕ . In terms of overall goodness-of-fit, the hybrid spatial model has on average a smaller DIC value (828.5) than the standard multilevel (1282.0), hybrid CAR (1095.0), and pure spatial (1094.7) models. However, this result requires cautious interpretation as it may be attributed to that the true data are simulated under a scenario characterized by strong spatial correlation and long effective range, resulting in bias against the standard multilevel model.

In the case of fully out-of-sample cross validation, the variation in parameter estimation across the four types of models is substantially reduced under Scenario 2 (weak spatial correlation and long effective range) and 3 (strong spatial correlation and short effective range). For example, the standard multilevel model has a smaller MSE for the estimate of σ_u^2 compared to the hybrid CAR model, although it still maintains the largest bias and lowest CR. Under Scenario 4 (weak spatial correlation and short effective range), the standard multilevel model performs nearly as well as the hybrid spatial model. On the other hand, the hybrid spatial model still outperforms the pure spatial model and the hybrid CAR model with respect to the estimates of σ_s^2 and ϕ under different scenarios. These findings hold regardless of whether the data come from partially or fully out-of-sample cross validation. Taken together, in the presence of non-ignorable within-neighborhood and distance-based spatial correlations, the main difference in parameter estimation is confined to the random effects variances. The hybrid spatial that appropriately adjusts for both sources of correlations leads overall to the best results considering all the aspects of estimation performance.

One difference between partially and fully out-of-sample cross validation comes from predictions for new observations. Model comparison in terms of predictive power is based on predictive match rate, that is, the percentage of correct predictions for the validation data

whose true outcome values are known *a priori* but assigned as missing at the modeling stage. There is almost no difference in the predictive power across the four different types of models when they are used to predict outcome values for new observations in existing neighborhoods (results not shown). In contrast, when predictions are made for new observations in new neighborhoods, the three spatial models are unsurprisingly superior to the standard multilevel model by a narrow margin. This is demonstrated in Figure 3 which shows the boxplots of the predictive match rates from the four types of logit models fitted to the fully out-of-sample cross validation data sets simulated from an exponential spatial correlation functions under different scenarios. The pure spatial model appears to have the highest predictive match rate, followed by the two hybrid models and then the standard multilevel model. However, the difference here is very minor by about 2–3% according to the median predictive match rates under Scenario 1, and it further drops as the spatial correlation weakens or has a shorter effective range as under Scenario 2–4. Similar results (not shown) are obtained when the data are simulated using a linear spatial correlation structure as specified in Equation (7), but the differences are less visible across the four models compared to simulations based upon an exponential function.

4.2. Robustness against Model Misspecification

Table 2 presents comparisons of parameter estimation for a common type of model misspecification where an important neighborhood-level predictor is left out, which is not uncommon in empirical studies. Parameter estimations of the intercept (β_0) and individual-level coefficient (β_1) are barely affected as compared to those under correct model specifications presented in Table 1, regardless of the relative strength or effective range of spatial correlation. Under Scenario 1, the estimations of random effects variances, σ_u^2 and σ_s^2 , as well as the distance-decay parameter, ϕ , are now subject to increased bias. Nonetheless, overall, the loss in the accuracy and precision of parameter estimation is quite limited in all but the hybrid CAR model, suggesting that most of the models are reasonably robust against misspecifying predictors in case of strong spatial correlation and long effective range. The results are quite mixed when the relative strength or effective range of the spatial correlation decreases (i.e. Scenarios 2–4). Nonetheless, a general pattern is that the hybrid spatial model has consistently better performance with respect to the random effects estimates than the other models, although it has more biased estimates of σ_u^2 under Scenarios 2 and 4.

The second type of model misspecification mimics the case when the true spatial correlation follows a Gaussian distance-decay function as specified in Equation (6) but an exponential function is mistakenly adopted to fit the spatial model. Performance measures for this type of model misspecification are presented in Table 3. Across different scenarios of the relative strength and effective range of spatial correlation, the most striking result is that the estimations for the spatial parameters σ_s^2 and ϕ are again severely biased, although the hybrid spatial model provides the best estimates for these two parameters. However, the hybrid spatial model also produces the most biased yet moderately accurate estimates of σ_u^2 compared to the standard multilevel and hybrid CAR models.

Figures 4 and 5 plot the predictive match rates for the two types of model misspecification, respectively. In both cases, the pure spatial and two hybrid models sustain their minor

predictive advantage over the standard multilevel model. In particular, the pure spatial model maintains about 2–3% higher predictive match rates compared to the standard multilevel model under Scenario 1, regardless of model misspecifications, although this very modest advantage further drops under Scenarios 2–4.

5. AN EMPIRICAL EXAMPLE

5.1. Child Mortality Data of 1880 Newark, NJ

This section briefly introduces a real data set of child mortality used to demonstrate the strength of applying spatial models to empirical research. The data come from the Urban Transition Historical GIS Project,² which uses historical census data to document the state of American cities from the end of the nineteenth century into the early twentieth century. All residents in 39 U.S. cities were geocoded according to their household addresses based on the full transcription of the 1880 Census of Population created by the Church of Jesus Christ of Latter-day Saints and made widely accessible through the North Atlantic Population Project (NAPP) at the Minnesota Population Center (MPC).³ The geocoded individual-level data provide a great opportunity for conducting a wide range of spatial analyses in historical American cities.

The analysis here draws on the data from Newark, the leading city in New Jersey in 1880, with a focus on three predominant ethnic groups: Irish, Germans and Yankees. Irish and Germans include both first- and second-generation immigrants, that is, those who were not born in the U.S., and those who were born in the U.S. but whose parents were not. Yankees are natives who were born white and whose parents were also native born.

Data on death records are drawn from a database available at the Department of State of New Jersey. The death records between June 1878 and June 1885, including death certificates, burial, reburial, transit, and disinterment permits, are recorded by the New Jersey Department of Health. New Jersey is well known for its accurate and complete reporting of vital statistics in the late 19th century. Among the entire 1880 U.S. Census death registration area, for example, New Jersey was one of only two states to provide reasonably accurate and nearly complete (more than 90 percent) registration of deaths (Galishoff 1988). Therefore, the death records between June 1878 and June 1885 in Newark can be considered fairly accurate and complete given the historical context.

The dependent variable is treated as binary for illustration purpose only, that is, whether a child was dead or not by age 5 (i.e. 1885). To compare multilevel and spatial models, neighborhoods are conceptualized and approximated by enumeration districts (EDs). The entire city of Newark in 1880 was divided into 71 EDs, each of which refers to an area that was assigned to an enumeration supervisor to count persons within the area and prepare census population schedules. Figure 6 illustrates a map of part of the city, where each dot represents a child's household location, the thin gray lines depict the street network, and the

²More detailed information regarding the Urban Transition Historical GIS Project and the data used in this study can be found at: <http://www.s4.brown.edu/>.

³For more information on the transcription of the 1880 Census of Population see: <http://www.nappdata.org/>.

thick black lines draw the boundaries of EDs. Children who lived in the same ED as recorded in the 1880 census were considered as nested within the same neighborhood.

A total number of 501 death records by June 1885 were identified among 6,762 individuals who were infants (of age 0 to 1 year old) in 1880 Newark. This analysis is focused on the 438 death records among 5,767 infants who were Irish, Germans or Yankees. These 5,767 cases resided in 5,558 households across 844 street blocks which in turn were nested within 71 enumeration districts (EDs). That is, on average, each ED consisted of about 12 street blocks, within each of which 7 children lived.

Individual-level predictors include child's gender and ethnicity. A child's ethnicity is determined by combining several variables, including race, place of birth, and parents' places of birth. For example, a white person who was born in Ireland (first-generation immigrant) or who was born in any state of the US but whose parents were born in Ireland (second-generation immigrant) is coded as an Irish immigrant. Household-level predictors include household header's age and socioeconomic status, and number of children in the household. Socioeconomic status is measured by a socioeconomic index (SEI) score coded by the Minnesota Population Center based on people's average education and earnings in each occupation as measured in 1950 and standardized to be a continuous value bounded between 0 and 100 with 0 indicating unemployed.

ED-level predictors include population density, median SEI of household heads, and an ethnic segregation measure known as Simpson's diversity index. For each ED, a score of Simpson's diversity index is calculated by assessing ethnic compositions across street blocks that lie within the ED's boundary as sub-area units. A larger score of Simpson's diversity index reflects more evenly distributed ethnic populations across all street blocks within an ED, whereas a smaller score indicates a tendency for some blocks to be predominated by one ethnic group while other blocks to be occupied by another group. The Simpson's diversity index is normalized to be bounded between 0 (indicating the least ethnic diversity —i.e. a single group dominates the neighborhood) and 1 (indicating greatest ethnic diversity —i.e. all groups appear in the neighborhood with equal proportions).

5.2. Results for Child Mortality Data

Table 4 presents the descriptive statistics of the dependent variable and predictors for child mortality in 1880 Newark, NJ. The death rate was highest for Irish children (11 percent) and lowest for German children (5 percent). The data were roughly equally stratified by gender. More than half of the children were Yankees, about one fifth were Irish, and the rest were Germans. The average age of household heads was about 35 years old with little variation by children's ethnicities. On average, Yankee children lived in the wealthiest households, whereas Irish children lived in the poorest households as measured by household head's SEI score. Yankee children also tended to have fewer siblings compared to Irish and German children. As for the neighborhood context, German children lived in the least ethnically diverse yet most crowded EDs compared to Irish and Yankee children. Yankee children lived in slightly wealthier EDs compared to Irish and German children.

Table 5 presents regression results from the standard multilevel, pure spatial, hybrid CAR, and hybrid spatial logit models where an exponential function for spatial correlation is employed. The DIC values indicate that the hybrid spatial model fits best to the data, followed by the pure spatial, hybrid CAR, and standard multilevel models in descending order. The estimated coefficients of the individual and household-level predictors are roughly the same across the four models. Boys were significantly more likely than girls to die. The risk of death was significantly greater for Irish children, but lower for German children than for Yankee children. Household head's SEI was negatively associated with child mortality. The main difference in the estimates of fixed-effects arises from the ED-level predictors. Specifically, ED-level ethnic composition as measured by the Simpsons' diversity index was estimated to be significantly related to child mortality in the standard multilevel model, but not so in the three spatial models. This discrepancy may be attributable to ignoring potential correlations between children living in nearby EDs and hence downwards-biased estimates of standard errors in the standard multilevel model, resulting in inflated statistical significance. This is partially supported by visually comparing the posterior estimates from the multilevel and the hybrid models plotted on a map (see Figure 7 and the discussion below). It is also worth noting that the point estimate Simpsons' diversity index is much larger from the standard multilevel model than that from the other models. Therefore, it is unclear whether the other models actually produced biased estimates for this parameter as compared to the standard multilevel model. This discrepancy is worth further investigation in future substantive research.

Turning to other parameters, the estimate of σ_w^2 , the variance of ED-level random effects, in the standard multilevel model is about 1.5–5.5 times as large as that in the two hybrid models. The estimate of σ_s^2 , the variance of between-ED spatial random effects, in the pure spatial model is roughly 2–3 times as large as that in the two hybrid models. These results imply that both the standard multilevel and pure spatial models may overestimate the variances of random effects as a result of failing to adjust for two sources of correlated data. In contrast, the two hybrid models are more likely to take into account both within- and between-ED correlations. Figure 7 depicts the posterior means of the child mortality risks that are attributable to non-spatial neighborhood effects neighborhood estimated from the standard multilevel model (on the top) and that are attributable to the spatial effects from the hybrid model (on the bottom), respectively. There was a clear spatially smoothed pattern of child mortality risks (Figure 7 bottom), with clusters of high-risk EDs surrounded by modest-risk EDs which in turn were adjacent to clusters of low-risk EDs. On the contrary, the risks due to non-spatial neighborhood effects were scattered without a clear spatial trend over the city (Figure 7 top).

Moreover, the estimate of ϕ , the spatial decay parameter, in the pure spatial model was slightly different from that in the hybrid spatial model. Figure 8 plots the spatial correlation against geographic distance based on the estimates of ϕ . Notice that the lower limit of the 95 percent credible interval of the pure spatial model nearly coincides with that of the hybrid spatial model and thus is almost invisible on the graph. The average spatial correlation dropped at a slightly faster rate as estimated in the hybrid spatial model than in the pure spatial model. The corresponding 95 percent credible interval was also slightly tighter from

the hybrid spatial model, reflecting more accurate estimation. The effective range, that is, the distance at which spatial correlation drops to 5 percent and can be considered gone, was estimated to be about 0.8 kilometers in the pure spatial model, about 160 meters greater than in the hybrid spatial model. This difference could have substantive implications for children whose daily activities were largely concentrated in their households and who lived in 1880 Newark, a predominantly walking city. On the other hand, the hybrid CAR model is less capable of inferring the geographic scale for the strength of spatial correlation since it is based on a pre-defined lattice-like neighboring structure.

6. DISCUSSION

The standard multilevel regressions which are widely used in neighborhood research typically ignore potential between-neighborhood correlation due to underlying spatial processes, and hence are subject to biased statistical inference. In contrast, spatial models make estimations and predictions over space by explicitly modeling the spatial correlations among observations in different locations. Through systematic comparisons of model estimations and predictions for binary outcomes using both simulation and empirical data, this study sheds new light on the relative strengths and weaknesses of spatial models as compared to the standard multilevel model, and is informative for future research on neighborhood and spatial effects. The spatial models are particularly applicable to studies of neighborhood effects that involve social and demographic processes across neighborhood boundaries within an urban system. A few examples in the literature include the spatial dynamics embedded in the effects of neighborhood-level collective efficacy on social control of children (Sampson, Morenoff, and Earls 1999), violent crimes (Sampson, Raudenbush, and Earls 1997), and birth weight (Morenoff 2003) in Chicago; the impact of neighborhood socioeconomic inequality on children's achievement in Los Angeles (Sastry and Pebley 2010); and the relation between level of areal flooding and return migration to New Orleans after Hurricane Katrina (Fussell, Sastry, and VanLandingham 2010).

Several important findings stand out from the simulation analysis conducted in this study. First, the standard multilevel and spatial models have similar performance with respect to estimating parameters of fixed effects as well as predicting new observations within existing neighborhoods. In other words, adjusting for either within- or between-neighborhood correlation alone is almost as good as adjusting for both when the main interest is to obtain good estimates of fixed effects predictors or to make partially-out-of sample predictions.

Second, only adjusting for one type of correlation does lead to biased and inaccurate estimates of the variances of random effects, be it in within-neighborhood or between-neighborhood correlations. This may have serious implications for research on neighborhood effects given the common practice of assessing neighborhood effects solely based on the parameter estimate of σ_u^2 (Diez-Roux 2004). The simulation analysis conducted here suggests that σ_u^2 can be overestimated if the presence of spatial correlation is not appropriately incorporated into the model. Therefore, neighborhood analysis embedded within a large city-wide system should be carried out with caution. Between-neighborhood correlation (due, for example, to spatial spillover effects) should be examined whenever spatial information is available before applying a standard multilevel model, at the risk of

overstating within-neighborhood correlation. On the other hand, solely relying on spatial correlation without recognizing the existence of non-spatial within-neighborhood correlation may lead to overestimating the variance of spatial random effects (σ_s^2). Moreover, a pure distance-based spatial model tends to overestimate the spatial decay parameter (ϕ), possibly as a result of ignoring within-neighborhood correlation, compared to the hybrid models. Hence, non-spatial neighborhood-level effects need to be taken into account in order to avoid making biased spatial inference.

Third, there is some evidence of minor advantage of spatially modeling between-neighborhood correlations becomes apparent when it comes to making predictions for observations from an unobserved neighborhood. In such a case, a spatial model has the capacity to borrow strength from observations in nearby neighborhoods that are observed and hence is able to improve predictive accuracy, compared to a standard multilevel model that does not pool data across neighborhoods. This advantage is quite limited (only about 2–3% higher predictive match rates) and decreases as the strength of spatial correlation decreases relative to that of within-neighborhood correlation. However, this potential merit of spatial modeling has begun to attract researchers who integrate spatial information into social-demographic survey data in order to avoid drawing biased statistical inferences based on the assumption of independence between spatial units (Borgoni and Billari 2003). Nonetheless, hybrid spatial models, be they lattice- or distance-based, are likely to sacrifice their predictive power to better fitting the observed sample data, whereas the pure spatial model sustains its relatively superior performance of fully out-of-sample prediction, despite model misspecifications.

Drawing on geocoded child mortality data of 1880 Newark, NJ, this research illustrates the advantages of applying a spatial modeling strategy in empirical demographic research. By taking into account both within- and between-neighborhood correlations, hybrid models are effective at appropriately adjusting for uncertainty in regression estimation and thus avoiding artificially inflated statistical significance. Moreover, the hybrid spatial models allow spatial random effects to be disentangled from non-spatial neighborhood random effects, facilitating the detection of distinct spatial patterns within the population of interest. Nevertheless, distance-based hybrid model can provide extra information about the geographic scale of spatial correlation, which is not readily available in a lattice-based hybrid model.

To conclude, this study demonstrates several merits of spatial modeling that call for researchers to reflect upon the assumption of between-neighborhood independence and the role of space in understanding contextual effects. The application of spatial modeling in neighborhood research may have traditionally been hindered by a lack of appropriate data with enough spatial information. As GIS techniques continue to advance and become more cost-effective, however, social science researchers have now made tremendous progress in incorporating spatial components into large-scale population survey data collection in both developed (e.g. Harris et al. 2009) and developing countries (e.g. Popkin et al. 2010). Future research should take the advantage of fast-growing spatial data collection and get equipped with spatial modeling as a common analytical strategy.

Spatial models are not free from limitations, however. First, computational costs remain high despite considerable improvement in computer hardware and the development of new algorithms. For example, model convergence is problematic if a complex distance-decay spatial correlation function such as the Gaussian is adopted. Likewise, exploratory simulation analysis (not shown) suggests that a hybrid spatial model for ordinal or count variables other than binary data very often runs into numerical problems and hence leads to either unstable estimates or failure to converge. Although maximum likelihood estimates or alike (e.g. pseudo maximum likelihood) other than Bayesian estimates have been developed, they are subject to difficulty in numerical integration for high-dimensional complex data as well. Nevertheless, computational costs for spatial models are likely to be considerably reduced with improvements in algorithms and computer hardware in the near future.

Second, an understated issue in the present study and yet an important one in empirical studies is how neighborhoods or areas are defined. Neighborhood effects involve a process in which place-based membership induces a shared exposure to certain factors either imposed exogenously such as public policy or arising endogenously such as neighborhood collective efficacy. Then the question becomes, what place effects are to be modeled (Arcaya et al. 2012) and therefore where should the boundary of the place be drawn? These questions cannot be answered without careful theoretical reasoning and comprehensive knowledge about the substantive subject under study, although some analytical tools can offer some assistance (Logan et al. 2011). Nevertheless, instead of precluding a spatial modeling approach, these limitations imply large room for fruitful future research.

Acknowledgments

The author thanks John R. Logan and the staff of the research initiative on Spatial Structures in the Social Sciences at Brown University for providing the historical GIS data used in this study. The author greatly acknowledges the support of the computational resources and services provided by the Center for Computation and Visualization, Brown University. The author also thanks Marcia C. Castro and participants at the 2012 Annual Meeting of the Population Association of America for helpful comments on the paper.

Funding: The historical GIS data collection used in this work was supported by the National Science Foundation [grant number 0647584] and the National Institutes of Health [grant number 1R01HD049493-01A2].

Biography

Hongwei Xu is a faculty research fellow at the Survey Research Center, Institute for Social Research, University of Michigan, and a research affiliate at the Population Studies and Training Center, Brown University. His methodology interests include spatial statistics, longitudinal and multilevel analysis, and Bayesian inference. His substantive research topics cover health inequalities, epidemiologic and nutrition transitions, child well-being, and the residential segregation of ethnic groups. He is currently working as a research team member on the Chinese Family Panel Studies, one of the largest longitudinal data collection projects in contemporary China.

REFERENCES

Akaike, Hirotugu. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974; 19:716–723.

- Arcaya, Mariana; Brewster, Mark; Zigler, Corwin M.; Subramanian, SV. Area variations in health: A spatial multilevel modeling approach. *Health & Place*. 2012; 18:824–831. [PubMed: 22522099]
- Banerjee, Sudipto; Gelfand, Alan E.; Carlin, Bradley P. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Press; 2004.
- Besag, Julian; York, Jeremy; Mollié, Annie. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 1991; 43:1–20.
- Borgoni, Riccardo; Billari, Francesco C. Bayesian spatial analysis of demographic survey data: An application to contraceptive use at first sexual intercourse. *Demographic Research*. 2003; 8:61–92.
- Brooks, Stephen P.; Gelman, Andrew. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*. 1998; 7:434–455.
- Browne, William; Goldstein, Harvey. MCMC Sampling for a Multilevel Model With Nonindependent Residuals Within and Between Cluster Units. *Journal of Educational and Behavioral Statistics*. 2010; 35:453–473.
- Browne, William J.; Goldstein, Harvey; Rasbash, Jon. Multiple membership multiple classification (MMM) models. *Statistical Modelling*. 2001; 1:103–124.
- Burton, Andrea; Altman, Douglas G.; Royston, Patrick; Holder, Roger L. The design of simulation studies in medical statistics. *Statistics In Medicine*. 2006; 25:4279–4292. [PubMed: 16947139]
- Chaix, Basile; Merlo, Juan; Chauvin, Pierre. 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*. 2005; 59:517–526. [PubMed: 15911650]
- Cohen, Jacqueline; Tita, George. Diffusion in Homicide: Exploring a General Method for Detecting Spatial Diffusion Processes. *Journal of Quantitative Criminology*. 1999; 15:451–493.
- Dietz, Robert D. The estimation of neighborhood effects in the social sciences: An interdisciplinary approach. *Social Science Research*. 2002; 31:539–575.
- Diez-Roux, Ana V. Multilevel Analysis in Public Health Research. *Annual Review of Public Health*. 2000; 21:171–192.
- Diez-Roux, Ana V. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Social Science & Medicine*. 2004; 58:1953–1960. [PubMed: 15020010]
- Diggle PJ, Tawn JA, Moyeed RA. Model-Based Geostatistics. *Applied Statistics*. 1998; 47:299–350.
- Diggle, Peter J.; Heagerty, Patrick; Liang, Kung-Yee; Zeger, Scott L. *Analysis of Longitudinal Data*. New York: Oxford University Press; 2002.
- Diggle, Peter J.; Ribeiro, Paulo J, Jr. Bayesian Inference in Gaussian Model-based Geostatistics. *Geographical & Environmental Modelling*. 2002; 6:129–146.
- Diggle, Peter J.; Ribeiro, Paulo J., Jr; Christensen, Ole F. An Introduction to Model-based Geostatistics. In: Moller, J., editor. *Spatial Statistics and Computational Methods*. New York: Springer-Verlag; 2003. p. 43-86.
- DiPrete, Thomas A.; Forristal, Jerry D. Multilevel Models: Methods and Substance. *Annual Review of Sociology*. 1994; 20:331–357.
- Flowerdew, Robin; Manley, David J.; Sabel, Clive E. Neighbourhood effects on health: Does it matter where you draw the boundaries? *Social Science & Medicine*. 2008; 66:1241–1255. [PubMed: 18177988]
- Fussell, Elizabeth; Sastry, Narayan; VanLandingham, Mark. Race, socioeconomic status, and return migration to New Orleans after Hurricane Katrina. *Population and Environment*. 2010; 31:20–42. [PubMed: 20440381]
- Galishoff, Stuart. *Newark: The Nation's Unhealthiest City, 1832–1895*. New Brunswick, NJ: Rutgers University Press; 1988.
- Gelfand, Alan E.; Latimer, Andrew; Wu, Shanshan; Silander, John A, Jr. Building statistical models to analyze species distributions. In: Clark, JS.; Gelfand, AE., editors. *Hierarchical Modelling for the Environmental Sciences: Statistical Methods and Applications*. New York: Oxford University Press; 2006. p. 77-97.
- Gelman, Andrew; Goegebeur, Yuri; Tuerlinckx, Francis; Van Mechelen, Iven. Diagnostic checks for discrete data regression models using posterior predictive simulations. *Applied Statistics*. 2000; 49:247–268.

- Gelman, Andrew; Hill, Jennifer. Data Analysis Using Regression and Multilevel/Hierarchical Models. New York: Cambridge University Press; 2007.
- Gelman, Andrew; Rubin, Donald B. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*. 1992; 7:457–472.
- Gilks, Walter R.; Richardson, Sylvia; Spiegelhalter, David. Markov Chain Monte Carlo in Practice. London: Chapman & Hall; 1996.
- Goldstein, Harvey. Multilevel Statistical Models. West Sussex, UK: Wiley; 2010.
- Guo, Jessica Y.; Bhat, Chandra R. Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*. 2007; 15:31–45.
- Harris KM, Halpern CT, Whitel E, Hussey J, Tabor J, Entzel P, Udry JR. The National Longitudinal Study of Adolescent Health: Research Design. 2009
- Hedström, Peter. Contagious Collectivities: On the Spatial Diffusion of Swedish Trade Unions, 1890–1940. *American Journal of Sociology*. 1994; 99:1157–1179.
- Henderson, Robin; Shimakura, Silvia; Gorst, David. Modeling Spatial Variation in Leukemia Survival Data. *Journal of the American Statistical Association*. 2002; 97:965–972.
- Kohavi, Ron. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. 1995; 2:1137–1143.
- Liu, Xuan; Wall, Melanie M.; Hodges, James S. Generalized spatial structural equation models. *Biostatistics*. 2005; 6:539–557. [PubMed: 15843593]
- Logan, John R.; Molotch, Harvey L. Urban Fortunes: The Political Economy of Place. Berkeley, CA: University of California Press; 1987.
- Logan, John R.; Spielman, Seth; Xu, Hongwei; Klein, Philip N. Identifying and Bounding Ethnic Neighborhoods. *Urban Geography*. 2011; 32:334–359. [PubMed: 24039327]
- Logan, John R.; Zhang, Weiwei; Xu, Hongwei. Applying Spatial Thinking in Social Science Research. *Geo Journal*. 2010; 75:15–27. [PubMed: 20431703]
- Lunn, David; Spiegelhalter, David; Thomas, Andrew; Best, Nicky. The BUGS project: Evolution, critique and future directions. *Statistics In Medicine*. 2009; 28:3049–3067. [PubMed: 19630097]
- Morenoff, Jeffrey D. Neighborhood mechanisms and the spatial dynamics of birth weight. *American Journal of Sociology*. 2003; 108:976–1017.
- Morenoff, Jeffrey D.; Sampson, Robert J. Violent Crime and the Spatial Dynamics of Neighborhood Transition: Chicago, 1970–1990. *Social Forces*. 1997; 76:31–64.
- Popkin, Barry M.; Du, Shufa; Zhai, Fengying; Zhang, Bing. Cohort Profile: The China Health and Nutrition Survey - monitoring and understanding socio-economic and health change in China, 1989–2011. *International Journal of Epidemiology*. 2010; 39:1435–1440. [PubMed: 19887509]
- Ribeiro, Paulo J.; Diggle, Peter J. geoR: A Package for Geostatistical Analysis. *R News*. 2001; 1:15–18.
- Riva, Mylene; Apparicio, Philippe; Gauvin, Lise; Brodeur, Jean-Marc. 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*. 2008; 7
- Sampson, Robert J.; Morenoff, Jeffrey D.; Earls, Felton. Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children. *American Sociological Review*. 1999; 64:633–660.
- Sampson, Robert J.; Morenoff, Jeffrey D.; Gannon-Rowley, Thomas. Assessing "Neighborhood Effects": Social Processes and New Directions in Research. *Annual Review of Sociology*. 2002; 28:443–478.
- Sampson, Robert J.; Raudenbush, Stephen W.; Earls, Felton. Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science*. 1997; 227:918–923. [PubMed: 9252316]
- Sastry, Narayan; Pebley, AnneR. Family and neighborhood sources of socioeconomic inequality in children's achievement. *Demography*. 2010; 47:777–800. [PubMed: 20879688]
- Snijders, Tom AB.; Bosker, Roel J. Modeled Variance in Two-Level Models. *Sociological Methods & Research*. 1994; 22:342–363.

- Spiegelhalter, David J.; Best, Nicola G.; Carlin, Bradley P.; Van Der Linde, Angelika. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society, Series B*. 2002; 64:583–639.
- Tanner, Martin A.; Wong, Wing Hung. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*. 1987; 82:528–540.
- Tatalovich, Zaria; Wilson, John P.; Milam, Joel E.; Jerrett, Michael; McConnell, Rob. Competing definitions of contextual environments. *International Journal of Health Geographics*. 2006; 5:55. [PubMed: 17156433]
- Thomas, Andrew; Best, Nicky; Lunn, David; Arnold, Richard; Spiegelhalter, David. *GeoBUGS User Manual (Version 1.2)*. 2004
- Tobler, Waldo R. A Computer Movieview Simulating Urban Growth in the Detroit Region. *Economic Geography*. 1970; 46:234–240.

High Mean Level of Individual Outcome

Medium Mean Level of Individual Outcome

Low Mean Level of Individual Outcome

Figure 1.
Hypothetical examples of multilevel (left) and spatial (right) models.

Figure 2.
A simulation of spatial random effects with an exponential function ($\sigma_s^2 = 3$, $\phi = 3$, and effective range = 1).

Figure 3.

Fully out-of-sample predictive match rates from logit models for 4-fold cross validation.

Note: Scenario 1 applies $\sigma_s^2=3, \sigma_u^2=3, \phi=3$; Scenario 2 applies $\sigma_s^2=1.5, \sigma_u^2=3, \phi=3$;

Scenario 3 applies $\sigma_s^2=3, \sigma_u^2=3, \phi=6$; Scenario 4 uses $\sigma_s^2=1.5, \sigma_u^2=3, \phi=6$. Both pure and hybrid spatial models use an exponential function for spatial correlation.

Figure 4.

Fully out-of-sample predictive match rates from mis-specified logit models for 4-fold cross validation: Mistakenly excluding neighborhood-level predictor.

Note: Scenario 1 applies $\sigma_s^2=3, \sigma_u^2=3, \phi=3$; Scenario 2 applies $\sigma_s^2=1.5, \sigma_u^2=3, \phi=3$;

Scenario 3 applies $\sigma_s^2=3, \sigma_u^2=3, \phi=6$; Scenario 4 uses $\sigma_s^2=1.5, \sigma_u^2=3, \phi=6$. Both pure and hybrid spatial models use an exponential function for spatial correlation.

Figure 5.

Fully out-of-sample predictive match rates from mis-specified logit models for 4-fold cross validation: Using an exponential function for Gaussian spatial correlation.

Note: Scenario 1 applies $\sigma_s^2=3, \sigma_u^2=3, \phi=3$; Scenario 2 applies $\sigma_s^2=1.5, \sigma_u^2=3, \phi=3$;

Scenario 3 applies $\sigma_s^2=1.5, \sigma_u^2=3, \phi=6$; Scenario 4 uses $\sigma_s^2=3, \sigma_u^2=3, \phi=6$. Both pure and hybrid spatial models use an exponential function for spatial correlation.

Figure 6.
Illustration of children's household locations, streets, and enumeration districts in a part of
1880 Newark, NJ

Figure 7.

Posterior means of neighborhood effects from the standard multilevel model (top) and that of spatial effects from the hybrid spatial model (bottom) fitted to the child mortality data of 1880 Newark, NJ.

Figure 8.
Estimates of spatial correlation from distance-based (exponential) pure and hybrid spatial logit models fitted to the child mortality data of 1880 Newark, NJ

Table 1

Performance measures for logit models fitted to 100 simulated data sets with 4-fold cross validation.

	Partially Out-of-Sample						Fully Out-of-Sample					
	Scenario 1						Scenario 2					
	TV	AE	PB	CR	MSE		TV	AE	PB	CR	MSE	
β_0	-0.5	-0.4	-22.4	0.5	0.6		-0.5	-0.4	-15.9	0.5	0.6	
Standard Multilevel							-0.5	-0.4	-15.9	0.5	0.6	1.1
Pure Spatial ^a	-0.5	-0.4	-25.8	0.9	0.7		-0.5	-0.4	-21.2	0.9	0.8	0.4
Hybrid CAR	-0.5	-0.4	-20.4	0.4	0.6		-0.5	-0.5	-7.5	0.5	0.8	0.3
Hybrid Spatial ^a	-0.5	-0.4	-24.7	0.9	1.0		-0.5	-0.3	-35.9	0.9	1.3	0.7
β_1	0.8	0.7	-12.5	0.8	0.0		0.8	0.7	-13.0	0.7	0.0	0.0
Standard Multilevel							0.8	0.7	-13.0	0.7	0.0	0.0
Pure Spatial	0.8	0.7	-12.5	0.8	0.0		0.8	0.7	-13.0	0.7	0.0	0.0
Hybrid CAR	0.8	0.7	-12.3	0.8	0.0		0.8	0.7	-12.1	0.9	0.0	0.0
Hybrid Spatial	0.8	0.7	-12.4	0.8	0.0		0.8	0.7	-12.8	0.7	0.0	0.0
γ_1	-0.5	-0.4	-13.1	1.0	0.1		-0.5	-0.5	-6.9	0.9	0.1	0.1
Standard Multilevel							-0.5	-0.5	-6.9	0.9	0.1	0.1
Pure Spatial	-0.5	-0.4	-12.3	1.0	0.1		-0.5	-0.5	-10.0	0.9	0.1	0.1
Hybrid CAR	-0.5	-0.4	-11.7	1.0	0.1		-0.5	-0.5	-4.1	1.0	0.2	0.1
Hybrid Spatial	-0.5	-0.4	-11.8	1.0	0.1		-0.5	-0.5	-8.9	0.9	0.1	0.1
σ_u^2	3.0	4.3	43.7	0.7	3.1		3.0	4.4	45.1	0.8	3.6	14.2
Standard Multilevel							3.0	4.4	45.1	0.8	3.6	14.2
Hybrid CAR	3.0	2.2	-25.6	0.9	1.5		3.0	3.2	5.1	1.0	15.5	1.0
Hybrid Spatial	3.0	2.2	-28.1	0.9	1.5		3.0	2.2	-27.6	0.9	1.6	1.4
Pure Spatial	3.0	6.0	98.9	0.6	15.3		3.0	6.5	116.4	0.6	22.9	15.0
σ_s^2												
Hybrid CAR	3.0	9.3	208.8	0.9	62.8		3.0	9.7	223.3	0.9	98.2	25.4
Hybrid Spatial	3.0	4.7	56.6	1.0	15.7		3.0	5.2	72.5	1.0	23.2	11.9
Pure Spatial	3.0	7.2	138.4	0.5	18.4		3.0	6.7	123.6	0.7	15.3	19.8
Hybrid Spatial	3.0	4.6	54.0	1.0	3.8		3.0	4.5	49.8	1.0	3.4	4.2

Fully Out-of-Sample											
Scenario 3						Scenario 4					
	TV	AE	PB	CR	MSE	TV	AE	PB	CR	MSE	
β_0	Standard Multilevel	-0.5	-0.4	-20.9	0.7	0.3	-0.5	-0.5	-8.7	0.9	0.1
	Pure Spatial ^a	-0.5	-0.4	-21.8	1.0	0.5	-0.5	-0.4	-11.1	1.0	0.2
	Hybrid CAR	-0.5	-0.4	-17.4	0.7	0.3	-0.5	-0.5	-7.0	0.8	0.2
	Hybrid Spatial ^a	-0.5	-0.4	-11.4	1.0	0.6	-0.5	-0.4	-14.5	1.0	0.5
β_1	Standard Multilevel	0.8	0.7	-13.8	0.7	0.0	0.8	0.7	-14.1	0.7	0.0
	Pure Spatial	0.8	0.7	-13.9	0.7	0.0	0.8	0.7	-14.1	0.7	0.0
	Hybrid CAR	0.8	0.7	-12.9	0.8	0.0	0.8	0.7	-13.5	0.8	0.0
	Hybrid Spatial	0.8	0.7	-13.6	0.7	0.0	0.8	0.7	-13.9	0.7	0.0
γ_1	Standard Multilevel	-0.5	-0.4	-11.7	0.9	0.1	-0.5	-0.5	-8.3	0.9	0.1
	Pure Spatial	-0.5	-0.4	-15.2	0.9	0.1	-0.5	-0.5	-9.7	0.9	0.1
	Hybrid CAR	-0.5	-0.5	-6.2	1.0	0.2	-0.5	-0.5	-8.5	1.0	0.1
	Hybrid Spatial	-0.5	-0.4	-12.8	1.0	0.1	-0.5	-0.5	-9.3	0.9	0.1
σ_u^2	Standard Multilevel	3.0	4.9	62.9	0.6	5.4	3.0	3.7	22.0	0.9	1.4
	Hybrid CAR	3.0	3.9	30.9	1.0	13.9	3.0	2.7	-11.3	1.0	0.8
	Hybrid Spatial	3.0	2.8	-6.7	1.0	1.2	3.0	2.3	-24.3	0.9	1.2
	Pure Spatial	3.0	6.8	127.9	0.4	22.4	1.5	5.1	240.3	0.0	15.7
σ_s^2	Hybrid CAR	3.0	10.1	237.7	1.0	116.5	1.5	6.1	304.1	1.0	31.5
	Hybrid Spatial	3.0	4.3	43.1	1.0	10.3	1.5	3.0	102.6	1.0	10.7
	Pure Spatial	6.0	7.2	20.0	1.0	2.2	6.0	7.4	22.9	1.0	2.6
	Hybrid Spatial	6.0	5.0	-17.4	1.0	1.7	6.0	5.0	-16.3	1.0	1.5

Note: TV = true value; AE = average estimate; PB = percentage bias; CR = coverage rate with 95% credible interval; MSE = mean square error.

^aExponential spatial correlation function.

Table 2

Performance measures for mis-specified logit models fitted to 100 simulated data sets with 4-fold fully out-of-sample cross validation: Mistakenly excluding neighborhood-level predictor.

Scenario 1										Scenario 2				
		TV	AE	PB	CR	MSE	TV	AE	PB	CR	MSE			
β_0	Standard Multilevel	-0.5	-0.4	-17.1	0.5	0.6	-0.5	-0.4	-17.3	0.6	0.3			
	Pure Spatial ^a	-0.5	-0.4	-18.8	0.9	1.0	-0.5	-0.4	-25.3	1.0	0.5			
	Hybrid CAR	-0.5	-0.4	-13.5	0.5	0.6	-0.5	-0.4	-16.1	0.6	0.3			
	Hybrid Spatial ^a	-0.5	-0.4	-21.1	0.9	1.4	-0.5	-0.4	-10.0	1.0	0.6			
	Standard Multilevel	0.8	0.7	-13.1	0.7	0.0	0.8	0.7	-13.8	0.7	0.0			
β_1	Pure Spatial	0.8	0.7	-13.2	0.7	0.0	0.8	0.7	-13.8	0.7	0.0			
	Hybrid CAR	0.8	0.7	-12.3	0.9	0.0	0.8	0.7	-13.2	0.8	0.0			
	Hybrid Spatial	0.8	0.7	-13.0	0.7	0.0	0.8	0.7	-13.6	0.7	0.0			
	Standard Multilevel	3.0	4.5	51.4	0.7	4.3	3.0	3.5	16.9	0.9	1.1			
	Hybrid CAR	3.0	3.0	0.3	1.0	3.8	3.0	2.5	-17.0	0.9	0.9			
σ_u^2	Hybrid Spatial	3.0	2.3	-24.0	0.9	1.6	3.0	2.2	-26.3	0.9	1.2			
	Pure Spatial	3.0	7.0	133.5	0.5	30.3	1.5	5.2	246.0	0.0	21.0			
	Hybrid CAR	3.0	9.6	221.5	0.9	85.6	1.5	5.8	287.4	0.9	26.4			
	Hybrid Spatial	3.0	5.4	80.7	1.0	25.2	1.5	3.0	103.3	1.0	8.3			
	Pure Spatial	3.0	6.7	124.3	0.7	15.6	3.0	7.4	146.2	0.4	20.1			
φ	Hybrid Spatial	3.0	4.5	50.0	1.0	3.6	3.0	4.8	59.9	1.0	3.9			
Scenario 3										Scenario 4				
		TV	AE	PB	CR	MSE	TV	AE	PB	CR	MSE			
β_0	Standard Multilevel	-0.5	-0.4	-20.2	0.7	0.3	-0.5	-0.5	-7.9	0.9	0.2			
	Pure Spatial ^a	-0.5	-0.4	-20.6	1.0	0.5	-0.5	-0.4	-15.3	1.0	0.2			
	Hybrid CAR	-0.5	-0.4	-17.7	0.7	0.3	-0.5	-0.5	-6.4	0.8	0.2			
	Hybrid Spatial ^a	-0.5	-0.5	-2.6	1.0	1.3	-0.5	-0.5	-2.0	1.0	0.5			

		Scenario 3					Scenario 4				
		TV	AE	PB	CR	MSE	TV	AE	PB	CR	MSE
β_1	Standard Multilevel	0.8	0.7	-13.9	0.7	0.0	0.8	0.7	-14.2	0.7	0.0
	Pure Spatial	0.8	0.7	-14.0	0.7	0.0	0.8	0.7	-14.2	0.7	0.0
	Hybrid CAR	0.8	0.7	-13.2	0.8	0.0	0.8	0.7	-13.7	0.8	0.0
	Hybrid Spatial	0.8	0.7	-13.8	0.7	0.0	0.8	0.7	-14.1	0.7	0.0
	Standard Multilevel	3.0	5.0	67.8	0.6	6.1	3.0	3.8	28.3	0.9	1.8
σ_u^2	Hybrid CAR	3.0	3.7	24.0	1.0	3.3	3.0	2.8	-5.4	1.0	1.1
	Hybrid Spatial	3.0	2.9	-3.4	1.0	1.2	3.0	2.4	-19.7	1.0	1.1
	Pure Spatial	3.0	7.0	133.5	0.3	23.1	1.5	5.4	258.8	0.0	18.5
σ_s^2	Hybrid CAR	3.0	9.5	218.1	0.9	78.4	1.5	6.3	321.5	0.9	36.7
	Hybrid Spatial	3.0	4.9	64.8	1.0	23.3	1.5	3.3	120.2	1.0	13.9
ϕ	Pure Spatial	6.0	7.3	20.9	1.0	2.4	6.0	7.4	24.0	1.0	2.9
	Hybrid Spatial	6.0	4.9	-18.5	1.0	2.1	6.0	5.0	-16.9	1.0	1.7
Note: TV = true value; AE = average estimate; PB = percentage bias; CR = coverage rate with 95% credible interval; MSE = mean square error.											
^a Exponential spatial correlation function.											

Table 3

Performance measures for mis-specified logit models fitted to 100 simulated data sets with 4-fold fully out-of-sample cross validation: Using an exponential function for Gaussian spatial correlation.

	Scenario 1					Scenario 2				
	TV	AE	PB	CR	MSE	TV	AE	PB	CR	MSE
β_0										
Standard Multilevel	-0.5	-0.3	-31.5	0.5	0.9	-0.5	-0.4	-23.1	0.5	0.5
Pure Spatial ^a	-0.5	-0.3	-37.5	0.9	1.1	-0.5	-0.4	-18.0	0.9	0.5
Hybrid CAR	-0.5	-0.4	-29.4	0.4	1.0	-0.5	-0.4	-21.7	0.4	0.5
Hybrid Spatial ^a	-0.5	-0.3	-43.7	0.9	1.7	-0.5	-0.3	-31.5	0.9	1.1
β_1										
Standard Multilevel	0.8	0.7	-12.9	0.8	0.0	0.8	0.7	-13.5	0.7	0.0
Pure Spatial	0.8	0.7	-12.9	0.7	0.0	0.8	0.7	-13.6	0.7	0.0
Hybrid CAR	0.8	0.7	-12.1	0.8	0.0	0.8	0.7	-13.0	0.8	0.0
Hybrid Spatial	0.8	0.7	-12.8	0.7	0.0	0.8	0.7	-13.4	0.7	0.0
γ_1										
Standard Multilevel	-0.5	-0.4	-15.5	1.0	0.1	-0.5	-0.5	-6.9	1.0	0.1
Pure Spatial	-0.5	-0.4	-18.5	1.0	0.1	-0.5	-0.5	-8.6	0.9	0.1
Hybrid CAR	-0.5	-0.4	-15.5	1.0	0.1	-0.5	-0.5	-5.2	1.0	0.1
Hybrid Spatial	-0.5	-0.4	-17.8	1.0	0.1	-0.5	-0.5	-7.2	1.0	0.1
σ_u^2										
Standard Multilevel	3.0	4.1	38.2	0.7	3.6	3.0	3.2	7.9	0.9	1.3
Hybrid CAR	3.0	2.1	-31.6	0.8	3.8	3.0	2.0	-33.0	0.8	3.2
Hybrid Spatial	3.0	1.6	-45.0	0.7	2.5	3.0	1.7	-42.9	0.7	2.2
Pure Spatial	3.0	7.2	141.2	0.7	40.4	1.5	4.8	223.1	0.1	18.4
σ_s^2										
Hybrid CAR	3.0	9.7	221.8	0.7	71.8	1.5	6.4	326.8	0.8	43.4
Hybrid Spatial	3.0	6.7	123.6	1.0	41.3	1.5	4.2	183.2	1.0	25.3
Pure Spatial	$\sqrt{3}$	5.8	235.5	0.6	19.8	$\sqrt{3}$	6.8	294.5	0.4	27.7
ϕ										
Hybrid Spatial	$\sqrt{3}$	3.7	111.7	1.0	5.5	$\sqrt{3}$	4.2	144.6	1.0	7.6

	Scenario 3					Scenario 4				
	TV	AE	PB	CR	MSE	TV	AE	PB	CR	MSE
β_0										
Standard Multilevel	-0.5	-0.3	-35.5	0.6	0.7	-0.5	-0.4	-23.9	0.6	0.3
Pure Spatial ^a	-0.5	-0.4	-17.7	0.9	1.0	-0.5	-0.4	-16.2	1.0	0.5
Hybrid CAR	-0.5	-0.3	-33.6	0.4	0.7	-0.5	-0.4	-22.8	0.6	0.4
Hybrid Spatial ^a	-0.5	-0.4	-17.1	1.0	1.5	-0.5	-0.4	-27.0	0.9	0.7
β_1										
Standard Multilevel	0.8	0.7	-13.3	0.7	0.0	0.8	0.7	-13.6	0.7	0.0
Pure Spatial	0.8	0.7	-13.3	0.7	0.0	0.8	0.7	-13.7	0.7	0.0
Hybrid CAR	0.8	0.7	-12.4	0.8	0.0	0.8	0.7	-13.1	0.8	0.0
Hybrid Spatial	0.8	0.7	-13.1	0.7	0.0	0.8	0.7	-13.5	0.7	0.0
γ_1										
Standard Multilevel	-0.5	-0.4	-18.7	0.9	0.1	-0.5	-0.4	-17.9	1.0	0.1
Pure Spatial	-0.5	-0.4	-19.2	0.9	0.1	-0.5	-0.4	-18.3	0.9	0.1
Hybrid CAR	-0.5	-0.4	-15.7	1.0	0.1	-0.5	-0.4	-16.6	1.0	0.1
Hybrid Spatial	-0.5	-0.4	-17.8	0.9	0.1	-0.5	-0.4	-17.5	0.9	0.1
σ_u^2										
Standard Multilevel	3.0	4.3	44.5	0.8	3.9	3.0	3.4	13.1	0.9	1.0
σ_s^2										
Hybrid CAR	3.0	2.2	-27.2	0.9	4.1	3.0	2.0	-32.1	0.8	1.5
Hybrid Spatial	3.0	1.4	-51.9	0.7	2.9	3.0	1.7	-42.6	0.8	2.2
Pure Spatial	3.0	7.6	152.4	0.7	39.5	1.5	5.2	246.4	0.1	20.5
ϕ										
Hybrid CAR	3.0	10.9	262.0	0.7	99.8	1.5	6.8	353.7	0.8	38.2
Hybrid Spatial	3.0	7.4	145.5	0.9	45.6	1.5	4.1	173.6	1.0	17.9
Pure Spatial	$\sqrt{6}$	5.5	125.7	0.8	12.1	$\sqrt{6}$	6.7	174.3	0.6	19.8
Hybrid Spatial	$\sqrt{6}$	3.8	55.0	1.0	3.2	$\sqrt{6}$	4.4	80.4	1.0	5.0

Note: TV = true value; AE = average estimate; PB = percentage bias; CR = coverage rate with 95% credible interval; MSE = mean square error.

^aExponential spatial correlation function.

Table 4

Descriptive statistics of the child mortality data in 1880 Newark, NJ

	Total (N = 5767)		Irish (N = 1071)		German (N = 1484)		Yankee (N = 3212)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Died by age 5	0.08	0.26	0.11	0.32	0.05	0.21	0.08	0.27
Individual Level								
Sex (male = 1, female = 0)	0.49	0.50	0.46	0.50	0.50	0.50	0.50	0.50
Ethnicity								
Irish (= 1, else = 0)	0.19	0.39	—	—	—	—	—	—
German (= 1, else = 0)	0.26	0.44	—	—	—	—	—	—
Yankee (= 1, else = 0)	0.56	0.50	—	—	—	—	—	—
Household Level								
Head's age	34.99	8.92	35.75	7.31	36.47	7.88	34.06	9.71
Head's socioeconomic index	27.57	20.01	20.76	16.64	27.15	18.64	30.03	21.09
Number of children	3.12	1.84	3.46	1.85	3.78	1.94	2.69	1.66
Enumeration District Level								
Simpson's diversity index	0.72	0.23	0.80	0.17	0.59	0.27	0.75	0.21
Population density (persons/km ²)	11497.27	7708.21	9975.34	6239.35	13507.99	9491.05	11075.74	7021.36
Median socioeconomic index	24.68	9.23	22.24	8.09	22.54	6.75	26.49	10.13

Table 5

Estimates of logit models of child mortality in 1880 Newark, NJ

	Standard Multilevel		Pure Spatial		Hybrid CAR		Hybrid Spatial	
	β	SE	β	SE	β	SE	β	SE
Individual								
Sex (ref = female)	0.218	0.094 *	0.213	0.091 *	0.217	0.098 *	0.202	0.088 *
Ethnicity (ref = Yankee)								
Irish	0.371	0.120 **	0.350	0.127 **	0.372	0.125 **	0.366	0.114 ***
German	-0.442	0.144 **	-0.429	0.147 **	-0.432	0.151 **	-0.455	0.151 **
Household								
Head's age	-0.002	0.006	-0.002	0.006	-0.001	0.006	-0.003	0.006
Head's socioeconomic index	-0.008	0.003 **	-0.008	0.003 **	-0.008	0.003 **	-0.008	0.003 **
Number of children	0.011	0.028	0.009	0.029	0.007	0.029	0.013	0.029
Enumeration District								
Simpson's diversity index	0.604	0.313 *	0.470	0.321	0.491	0.338	0.459	0.365
Population density	0.050	0.069	0.057	0.081	0.041	0.078	0.053	0.076
Median socioeconomic index	-0.001	0.007	0.002	0.008	0.000	0.007	0.000	0.007
Random Effects								
σ_{η}^2	—	—	0.488	0.380	0.174	0.174	0.244	0.16
σ_u^2	0.288	0.08	—	—	0.052	0.046	0.19	0.11
Spatial Correlation								
ϕ	—	—	3.754	3.019	—	—	4.711	2.752
$3/\phi$	—	—	0.799	—	—	—	0.637	—
Constant	-2.857	0.349 ***	-2.826	0.364	-2.355	1.012	-2.704	0.479 ***
DIC	3455	—	3429	—	3435	—	3426	—

Notes: (1) Each model is fitted by running 3 MCMC chains, each of which runs 80,000 iterations with the first half discarded as the burn-in; the posterior distributions are summarized from the 1,000 iterations after thinning.

(2) β and SE are the means and standard deviations of the posterior distributions.

(3) * $p < .05$; ** $p < .01$; *** $p < .001$.