A SYNTHESIS OF SPATIAL MODELS FOR MULTIVARIATE COUNT RESPONSES

Yiyi Wang
Assistant Professor, Civil Engineering Department
Montana State University

Kara Kockelman
Professor, Department of Civil, Architectural, and Environmental Engineering
The University of Texas at Austin

Amir Jamali
Graduate Research Assistant, Civil Engineering Department
Montana State University

Pre-Print of Chapter published in *Regional Research Frontiers Vol 2: Methodological Advances, Regional Systems Modeling and Open Sciences* (2017), (Eds. Randall Jackson & Peter Schaeffer, by Springer Cham) pp. 221-237, Available in published form at https://link.springer.com/content/pdf/10.1007%2F978-3-319-50590-9_14.pdf

Abstract

This chapter provides a synthesis of spatial data mining models for analyzing multivariate count responses. Geo-referenced multivariate count responses are common in regional science (e.g., registered vehicle counts by body type and firm/job counts by industry type), but are computationally difficult to model - especially when sample size is large. This chapter synthesizes relevant research and offers lessons learned and best practices for future research.

Key words: spatial econometric model, spatial autocorrelation, multivariate response, Bayesian estimation body type

Introduction

Spatial data are central to regional science applications and many other disciplines. Location attributes for each observation reveal where events occur or other information (pollution levels [Goodkind et al. 2014], land values [Du and Mulley 2012], and crimes [Levine, 2009]) exists, often at fine spatial resolution. There are three types of spatial data: geostatistical data, areal or lattice data, and point data.

- Geostatistical data are innate to the landscape or environment (such as soil mineral levels, rainfall, and pollutant levels) and span continuously over space. Given their continuous nature, such variables need to be collected by sampling at different locations (Deutsch and Journel 1997). The goal of geostatistical analysis is to predict values at unknown locations using sampled/observed values. For this purpose, kriging is often used: it spatially interpolates unknown values using observations nearby (Krige 1951).
- Areal or lattice data are observed at certain geographic units (e.g., vehicle registration data across counties and land use changes across parcels). These geographic units divide up the study area into small tiles (tessellations) like census tracts. The goal of areal data analysis is usually to detect and explain spatial patterns, as opposed to predicting unknown values since there is typically no gap in the study area of interest. Areal data are usually analyzed by spatial econometric methods (LeSage and Pace 2009; Anselin 2010).
- Point data note the location of many specific occurrences like crashes or species sightings over a period of time. "Hot Spot" analysis is often used to identify clustering patterns of these points (Lu 2000). An array of metrics can be used to portray the magnitude of clusters, like Moran's I, Geary C's Location Quotient, and the nearest neighbor index (NNI). Point data can be converted to areal data by tessellating the study area into zones and aggregating the points at each zone.

Motivations for Spatial Models

This chapter focuses on spatial models for analyzing areal data, in a multivariate count format (like vehicle ownership across census tracts, number of crimes across zones, and patent applications across counties). Spatial models are attractive for two reasons that are rooted in geospatial theory: spatial dependence and spatial heterogeneity.

Spatial dependence (autocorrelation) describes correlations across the same variable observed at different locations (zones). A positive spatial autocorrelation implies clustering, so values observed at nearby locations are more similar than values observed at distant locations. A negative spatial autocorrelation portrays a dispersed pattern, in which a value at one location tends to be surrounded by dissimilar values (for the same variable). Spatial heterogeneity is defined as uneven distribution of a variable over space (Vinatier et al. 2014). Spatial heterogeneity arises due to structural instability: each zone/location subscribes to a different process to generate the variable of interest. Spatial heterogeneity can be expressed in an analytical model either as heteroscedastic (non-constant) error variance or regression coefficients that vary across observational units (Anselin 2001). Simoes and Natario (2016) provide a summary of statistical tests to detect spatial heterogeneity.

Conventional econometric models do not work for data that exhibit spatial dependence and/or heterogeneity. These models assume that the error terms are distributed normal (Gaussian), retain the same variance (which violates spatial heterogeneity), and are independent across

observations (which conflicts with spatial dependence). To address spatial dependence, models that recognize correlations (such as spatial autoregressive models) have been rather effective in various contexts, like crash and crime prediction (Levine et al. 1995a, 1995b; Miaou et al. 2003; Wang and Kockelman 2013), home prices (Case et al. 2003), land use dynamics (Chakir and Parent 2009; Wang and Kockelman 2009; Wang et al. 2012), and technology innovations (LeSage and Pace 2010). To tackle spatial heterogeneity, geographically weighted regression (GWR) is regularly used through locally estimating coefficients, rendering a contextual layer of coefficient estimates that vary over space. Examples of GWR span many fields, such as ecology, wealth and epidemics (Platt 2004, Ognev-Himmelberger et al. 2009, Atkinson et al. 2003, and Nagaya et al. 23 2010), traffic count and crash count predictions across road networks (Zhao and Park 2004 and Hadayeghi et al. 2010), and land use (Páez 2006; Wang et al. 2011).

Geo-Referenced Multivariate Count Data

One form of areal/lattice data is geo-referenced count data, data that take on non-negative integer values and record the number of items or events in zones of interest (e.g., number of vehicles owned across zones, crime counts across block groups, and crash counts by intersection and/or road segment). For a generic count variable, multiple levels of that variable are often observed: for example, number of vehicles by fuel economy category or number of crimes by type. These are multivariate count data. It is often of interest to gauge correlations among the different levels of a count (response) variable in addition to incorporating spatial dependence and/or heterogeneity across locations. The correlations reveal interactions among different levels of the response variable.

This chapter provides a synthesis of spatial models for analyzing count responses that have location attributes. The synthesis begins with a discussion of univariate count responses before presenting methods for multivariate settings.

Spatial Models for Univariate Count Data

Techniques for analyzing spatial count data broadly diverge depending on the type of spatial interaction one wishes to control for. As noted earlier, there are two types of spatial interactions: spatial heterogeneity and spatial dependence. GWR seeks to address spatial interactions shown as contextual variations in coefficient estimates over space (i.e., spatial heterogeneity). Hadayeghi et al. (2010) developed a GWR-Poisson model to explain traffic crashes using transportation planning factors while controlling for spatial variations across zones. For each zone, a weighted Poisson regression model was estimated using the part of data set observed in that zone's neighborhood. Weights are assigned to all neighbors, to reflect their importance in predicting counts in the zone of interest. The weights fall as the (straight-line or network-based) distance between zones increases.

For spatial dependence, many methods exist for analyzing univariate count data. They generally fall into three categories, as follows.

Log-linear Spatial Models

Standard spatial models (e.g., spatial autoregressive [SAR] models or spatial error models [SEM], LeSage and Pace 2009) were developed for data generated from a Gaussian process, in which the response variable takes on a continuous form. While not inherently designed to analyze count data, these models are sometimes used for analyzing count responses that are high in magnitude (e.g., hourly traffic volumes and employment counts). To apply these models in a

count response setting, the count variable is artificially transformed into a quasi-continuous variable. A count variable (e.g., species abundance or counts) is typically normalized by an exposure term so that the resulting variable represents the rate at which things happen (e.g., species abundance per square mile or an approximation of crime counts per capita). Then, the rate variable is log-transformed, to produce a new response variable. The log transformation is important because it allows for the possibility of negative predictions. Examples include Weir et al's. (2009) study on pedestrian crashes across San Francisco census tracts and Aufhauser and Fischer's (1985) study on migration patterns.

However, the log transformation will not work when low or zero counts exist, since their logarithms are mathematically ill-defined. In addition, a Gaussian process falls short of describing discrete events (e.g., crime or traffic crashes) that have low counts (rates), making it more attractive to use a discrete random process (e.g., Poisson). Two general approaches for discrete data analysis exist: these are conditional autoregressive (CAR) Poisson models and autoregressive Poisson models. Their difference lies in where spatial autocorrelation occurs: across the error terms (as in the CAR-Poisson) or the response terms (in the autoregressive-Poisson). *Conditional Autoregressive (CAR) Poisson Models*

A CAR-Poisson model assumes that the count variable follows a Poisson process: $y_i \sim$ Poisson (λ_i), where y_i represents the number of events observed in zone i, and λ_i denotes the expected/mean count for that zone. The expected mean relates to the explanatory variables (x_i), their coefficients, and an exposure term (E): $\lambda_i = E^{\alpha} \cdot \exp(x_i'\beta \quad + \gamma_i)$. The nuisance term, γ_i, represents noise or uncertainty that is unexplained by the control variables and is assumed to follow a CAR specification.

CAR specifications apparently were first used by Besag (1975), and are mostly estimated using Bayesian methods. A CAR model is built from a series of conditional distributions,[1] as shown in Equation 1 (Cressie 1993):

$$\gamma_i|\gamma_{\neq i} \sim N\left[\mu_i + \sum_{j=1}^{n} c_{ij}\left(\gamma_j - \mu_j\right),\ \sigma_i^2\right] \qquad (1)$$

where γ_i indicates the spatially autocorrelated variable (e.g., spatial random effects centered onzero, or a response variable -- like traffic flows or household incomes), γ_{-i} denotes such variables at neighboring locations (other than location i), μ_i is the expected/mean value of γ_i

[1] These conditional distributions lead to a multivariate normal (MVN) joint distribution of the spatially correlated variables (shown in Equation 2), based on the factorization theorem (Besag 1975).

$$\boldsymbol{\gamma} \sim MVN_n[\boldsymbol{\mu}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}] \qquad (2)$$

where the column vector $\boldsymbol{\gamma}$ is a stacked version of the n γ_i's (as is the vector $\boldsymbol{\mu}$), $\mathbf{I}$ is an identity matrix, $\mathbf{C}$ is an n by n weight matrix (defined by site contiguity or inter-observation distances), with $\mathbf{C} = [c_{ij}]$, and $\mathbf{M}$ is a diagonal matrix, with $\mathbf{M}_{ii} = \sigma_i^2$. This joint distribution is used along with the likelihood function of the data set to implement the Gibbs sampler to estimate the posterior distributions of all parameters. Note that the Equations (1) and (2) are often referred to as a Markov random field (MRF) because of the way they are derived: achieving a closed-form joint distribution by first specifying a set of conditional distributions (Banerjee et al. 2004).

The validity of the MVN distribution shown in Equation 2 requires that its covariance matrix, $(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}$, be symmetric and positive-definite (like any covariance matrix must), thereby necessitating certain constraints on the forms of the matrices $\mathbf{C}$ and $\mathbf{M}$. For example, one may let $\mathbf{C} = \rho\mathbf{W}$ and $\sigma_i^2 = \frac{\sigma^2}{w_{i+}}$, where ρ is referred to as the spatial autocorrelation coefficient, $\mathbf{W}$ is a row-standardized weight matrix (i.e., $\mathbf{W} = [w_{ij}^*]$ and $w_{ij}^* = \frac{w_{ij}}{w_{i+}}$), and w_{i+} is the i^{th} row sum of $\mathbf{W}$.

(i.e., $E(\gamma_i) = \mu_i$) and assumed to be zero, σ_i^2 is the conditional variance, and c_{ij} are weights (either known or unknown) describing the proximity or closeness between locations *i* and *j*.

The CAR specification permits contiguity and distance-based weight matrices, but precludes the K^{th}-nearest-neighbor weighting scheme because such weights violate the symmetry condition. First-order contiguity weights are defined such that $w_{ij} = 1$ if *i* and *j* share a common border (else $w_{ij} = 0$), and **W**'s diagonal elements are all zeros by construction (Cressie 1991). This type of CAR model is called a *proper* CAR model[2], and is commonly estimated using Bayesian techniques in the open-source WinBUGS software package (Spiegelhalter 2003), where "BUGS" stands for Bayesian inference Using Gibbs Sampling.

Spatial Autoregressive Poisson Models

While the CAR-Poisson model captures spatial dependence in error terms, SAR-type models describe spatial dependence in response variables. Lagrange multiplier tests can be used to discern which type of spatial dependence prevails in a spatial data set (whether spatial dependence occurs across the error terms or the responses) (Simoes and Natario, 2016). Intuitively, a spatially-lagged error term represents subtle spatial dependence due to missing variables that trend in space, whereas a spatially-lagged response term implies more direct spatial interactions in which the response observed at one zone is in part predicted by its neighbors' values in addition to its own covariates.

Cressie (1991) introduced the auto-Poisson model, in reference to models in which the mean rate, $\boldsymbol{\lambda}$, involves autocorrelated response variables, i.e., $\boldsymbol{\lambda}=\exp(\mathbf{X}\boldsymbol{\beta}+\rho\mathbf{W}\boldsymbol{y})$. More recently, Griffith (2000) and Chun (2008) developed a Poisson-based spatial filtering approach to estimate auto-Poisson models. However, these types of Poisson models permit only negative autocorrelation, an unwanted result arising from the peculiar way spatial autocorrelation enters the specification, as shown in the following equation: $\boldsymbol{\lambda}=\exp(\mathbf{X}\boldsymbol{\beta}+\rho\mathbf{W}\boldsymbol{y})$, where $\boldsymbol{\lambda}$ denotes a vector of expected mean rates, **X** is an *n* by *k* covariate matrix, $\boldsymbol{\beta}$ is a *k* by 1 vector of unknown coefficients, $\boldsymbol{y}$ represents a vector of observed (count) responses, **W** an *n* by *n* weight matrix, and ρ the spatial autocorrelation coefficient. In addition, the joint likelihood function under an auto-Poisson assumption requires a non-closed-form solution for the normalizing constant (in order for the joint likelihood function under the auto-Poisson specification to be proper, or integrate to1), which impedes successful estimation (Griffith 2000).

Liesenfeld et al. (2015) developed a new method to estimate spatial models for a wide range of non-Gaussian response variables including discrete choices, count, and other limited dependent variables (e.g., truncated, censored, or self-selected). This method combined Efficient Important Sampling (EIS) and sparse matrix algorithms to achieve accurate estimation of the likelihood function associated with spatially-interacted data and can handle a large number of observations. Liesenfeld et al. (2015) provided two such demonstrations: a spatial probit model for understanding U.S. voters' decisions in the 1996 presidential election, and a spatial count model for anticipating the prevalence of start-up companies across 3,110 U.S. counties.

For count responses, the model is formulated as:

[2] This is not the "intrinsic" CAR model, because the latter does not have a spatial autocorrelation coefficient, ρ, which measures the overall strength of spatial interactions. Due to the absence of the spatial autocorrelation coefficient, its joint distribution is improper or unbounded in the sample space (Gelfand and Vounatsou 2003).

$$f\ (y\ |\lambda, X) = \prod_{i=1}^{n} f\ (y_i\ |\lambda_i) \quad (3)$$

$$ln(\lambda)|X \sim MVN(m, H^{-1}) \quad (4)$$

where, y is the response variable (e.g., counts), $f\ (\cdot)$ is the likelihood function (e.g., $f\ (y) = \frac{\lambda^y e^{-\lambda}}{y!}$ for a Poisson process), λ_i is a latent variable measuring the expected mean counts, *i* is an index for observation unit. The latent vector, $\ln(\lambda)$, follows a multivariate normal distribution centered at *m* with a variance-covariance matrix, H^{-1}, i.e., the inverse of a Hessian matrix, *H.* When a direct spatial interaction is anticipated among neighbors, the latent variable at location *i* is influenced by the latent variables observed at its neighbors: mathematically, $\ln(\lambda) = \rho W \ln(\lambda) + X\beta + \varepsilon$. Under this construct, the mean (m) and the Hessian matrix are defined by $m = (I - \rho W)^{-1} X\beta$ and $H = (1/\ \sigma^2)\ (I_n - \rho W)'\ (I_n - \rho W)$. The rest of the parameters (ρ and W) are as defined previously.

Spatial Models for Multivariate Count Data

While univariate count models address spatial dependence for a single outcome across zones, many empirical studies are interested in gauging the interactions among multiple outcomes while controlling for spatial effects. For example, the prevalence of one disease can coincidentally affect other diseases due to shared risk factors; the growth rate of new business establishment from one industry can correlate with those of other industries in nearby areas as a result of knowledge flows and transportation accessibility; and traffic crashes often show correlations among different severity levels because of shared influence of certain infrastructure or environmental factors that are latent/unobserved in the data. To control for these interactions among more than one outcome, multivariate (MV) count models are used to simultaneously anticipate the prevalence of multiple levels of outcomes while controlling for spatial effects.

In general, four methods exist for predicting MV count data over space in the literature. Table 1 summarizes research studies that utilized (spatial) MV count models in light of sample size, estimation method, statistical tools used, and model specifications.

Multivariate Conditional Autoregressive (MCAR) Models

The conditional autoregressive (CAR) model is the most commonly used method to handle spatial count data (e.g., Jin et al. 2005; Kramer and Williamson 2013; Barua et al. 2014; Boulieri et al. 2016). Its popularity is fueled by open-source software such as WinBUGS and its twin package OpenBUGS (Spiegelhalter, et al. 2003), which code and estimate the CAR specification and its extensions (e.g., a time-space CAR model and moving-average models) with hierarchical Bayesian methods.

A multivariate CAR structure builds upon the univariate CAR-Poisson structure noted earlier and was enhanced by studies in genome analysis (Gelfand and Vounatsou 2003), disease mapping (Jin et al. 2005), traffic safety (Wang and Kockelman 2013; Aguero-Valverde et al. 2016), alternative-fuel vehicles (Chen et al. 2013 and Bansal et al. 2014), and location decisions of new business establishment (Wang and Kockelman, 2016).

The CAR structure defines the spatial random term for response type *k* observed in zone *i* (ϕ_{ik}) by a multivariate normal distribution. For an example involving two levels of responses, $\boldsymbol{\phi}_2 \sim N(\mathbf{0}, [(\mathbf{D} - \alpha_2 \mathbf{W})\tau_2]^{-1})$ and $\boldsymbol{\phi}_1 | \boldsymbol{\phi}_2 \sim N(\mathbf{A}\boldsymbol{\phi}_2,\ [(\mathbf{D} - \alpha_1 \mathbf{W})\tau_1]^{-1})$, where τ_1 and τ_2 scale up or down the variance-covariance matrices; α_1 and α_2 measure the strength of spatial

dependence; **W** is the spatial weight matrix (defined by contiguity or distance, though the former is more common in empirical studies, thanks to the computational benefits of sparse matrices); and **D** is a diagonal matrix with the i^{th} diagonal element denoting the i^{th} row sum of the weight matrix **W**. More details are deferred to Wang and Kockelman (2013) for a two-level response setting and Bansal et al. (2014), Gelfand and Vounatsou (2003), and Aguero-Valverde et al. (2016) for response variables involving three or more levels.

Finite Mixture Models with Spatial Dependence

A standard finite mixture model provides a flexible alternative to analyze heterogeneous data and is typically estimated by the expectation-maximization (EM) algorithm (Gupta and Chen 2010). In a finite mixture model, the probability density function for a population (data) is expressed by a weighted average of the distribution functions of its sub-populations:

$$p(y|\Theta) = w_1 f_1(y|\theta_1) + w_2 f_2(y|\theta_2) + \cdots + w_K f_K(y|\theta_K) \quad (7)$$

where $\Theta = (\theta_1, \theta_2, \cdots, \theta_K ; w_1, w_2, \cdots, w_K)$ represents the parameter space; the weights are positive and sum to numeral one; and $f(\cdot)$ represents a distribution function (e.g., Poisson distribution with a latent parameter, λ_k, to measure the mean/expected level for a sub-population, if *y* is count). The model captures heterogeneity by compartmentalizing the probability density function of the population into discrete components associated with the sub-populations (Park 2010).

For spatial data, these discrete components can serve as proxies for the geographical clusters that exhibit unique trends or coefficients, hence controlling for area-specific heterogeneity (Alfo et al. 2009). Alfo et al. (2009) extended a standard finite mixture to control for spatial dependence within each cluster using the convolution method (also known as the Besag-York-Mollie [BYM] model, Besag et al. 1991) and the correlations among two levels of outcomes (e.g., two diseases). Specifically, the log-transformed mean is decomposed into three parts:

$$\log(\lambda_{k1}) = \alpha_1 + \mu_k + \beta_{k1} \quad (8)$$

$$\log(\lambda_{k2}) = \alpha_2 + \mu_k/\delta + \beta_{k2} \quad (9)$$

where, α_1 and α_2 are constant terms representing the base-line risks associated with each disease, μ_k represents the shared factors that influence both outcomes, and β_{k1} and β_{k2} represent factors specific to each outcome. In addition to area-specific heterogeneity (a fortuitous property of all finite mixture-type models), this model specification also allows for spatial dependence across clusters by imposing a CAR structure to the three random terms, μ_k, β_{k1}, and β_{k2}.

The model was applied to estimate the prevalence of two heart diseases across 375 boroughs in Italy's Lazio region (Alfo et al. 2009), among other applications in health geography (see, e.g., Anderson et al. 2014). While the finite mixture models can define clusters in a meaningful way, the models can incur excessive computation time and are considered a special type of the generalized MCAR models (Alfo et al. 2009).

Generalized Ordered-Response Models

Some researchers have modeled spatial count data from an *ordered response* perspective that is rooted in utility-maximization choice theory. For example, in the context of intersection pedestrian crashes, Castro et al. (2012) utilized a continuous latent variable to proxy for traffic crash propensity and defined cut-off values to divide the latent variable into mutually exclusive

intervals, with each interval representing a certain level of crash frequency. The model was cast in an ordered probit setting and estimated by a composite marginal likelihood approach.

Bhat et al. (2014) enhanced the model by permitting multivariate correlations through a multinomial probit (MNP) kernel. A MNP model is traditionally used in consumer choice or decision science to anticipate the influences of external variables on a person's choices (e.g., voting decision, vehicle purchase choice, etc.). In the context of multivariate count data modeling, each choice alternative can be used to represent each level of outcome. This method takes advantage of the quasi-concave property of the utility function and associated computational benefits. The model was estimated using the maximum composite marginal likelihood (MACML) approach (Bhat 2011).

Spatiotemporal Models

Aldor-Noiman et al. (2013) accounted for spatial and temporal dependencies in modeling weekly counts of different violent crimes across 188 Washington D.C. census tracts. Four crime types were analyzed simultaneously: rape, robbery, arson, and aggravated assault. The data present two challenges: low counts and irregular spatial structure. In the study area, two disjointed zones have crime rates that are correlated and nearby zones have opposite crime rates (due to heterogeneous demographics and natural boundary), diverging from a regular spatial data with clear spatial clustering. An integer-valued first-order autoregressive process, INAR(1), was used to capture temporal correlations among weekly crime rates. The use of INAR(1) is innovative because it incorporates two latent factors: a random term for seasonal effects and a zone-specific rate function that carries spatial dependence through a Dirichlet prior. A nonparametric Bayesian approach was used to estimate the multivariate Poisson-INAR(1) model, coupled with multiple shrinkage to handle the large sample size. "Bayesian nonparameteric methods have previously been studied as tools for data-driven clustering analysis" (Aldor-Noiman 2013, p. 4) and appear to be as an effective way to analyze multiple correlated low-count time series (e.g., wild fires and earthquakes). The Dirichlet process also offers advantages by presenting a sparse neighborhood structure, similar to what a sparse spatial weight matrix functions in a Bayesian parametric setting.

Conclusions

This chapter describes the various spatial models that have been used to analyze univariate and multivariate count responses with location attributes. Two types of spatial effects are generally considered: spatial dependence (i.e., interactions among neighbors directly through spatially correlated response terms or indirectly through spatially lagged nuisance terms) and spatial heterogeneity (to describe contextual differences via spatially variable coefficient values).

For univariate count data, many spatial models exist, including a CAR model to explain spatial dependence in the error terms, a Poisson autoregressive model to convey more direct influence among neighbors through the response terms, and a GWR-Poisson model to allow coefficients that vary across locations. Goodchild and Haining (2004) suggested that the CAR model best applies to regions having more "local" spatial effects, like first-order-neighbor influence, whereas other spatial stochastic processes (which include the SAR and spatial error models [SEMs]) are more suitable for situations with higher-order dependencies, and thus more "global" spatial effects or relationships/interactions.

For multivariate count data, spatial effects enter the models chiefly through CAR-type interactions across error terms. The multivariate CAR structure is the most common approach to analyze such data due in part to the wide usage of open-source statistical software. However, such models only describe spatial interactions across the error terms and fall short when a more direct representation of spatial interaction is desired. By comparison, generalized ordered response (GOR) models (Bhat 2011), the spatial autoregressive Poisson model (Liesenfeld et al. 2015), and the Poisson mixture models (e.g., Schmidt and Rodriguez 2010) offer more flexible specifications: e.g., the spatial autoregressive Poisson models allow for direct spatial interactions of a variety of limited dependent variables, and the GOR models and the Poisson mixture models permit both negative and positive correlations among response levels. Future research should consider testing among these methods with respect to prediction accuracy, transferability, and computation. Efforts could also be spent to explore new ways to expand the computation of multivariate count models as large-scale spatial data (e.g., GPS traces and naturalistic driving data) become more regularly recorded and used in geography, transportation, and regional science.

The future of spatial multivariate count modeling presents both challenges and opportunities. On challenges, a foremost one is small sample size as seen in the moderate number of observation units used in many studies reviewed. With the advent of crowdsourcing and voluntary geographic information, comes the need for analytical tools that can handle thousands of data points made over a large geography (e.g., pavement cracks observed across a road network, public opinions on designs or prototypes of a commercial product [Brabham 2008], and GPS traces of trips made by millions of households across a region) while portraying complex spatial (and temporal) interactions. The most common tool used so far is OpenBUGS, an open-source software that implements a number of complex spatial and time-series models through Bayesian MCMC methods (e.g., Gibbs sampling and Metropolis-Hastings algorithms). It is a variation of WinBUGS, which can also handle spatial models but is restricted to only one sampling method (Gibbs sampling).

Another challenge relates to computing issues (e.g., long run time and convergence) that complex models frequently encounter. While models involving moderate sample size (e.g., hundreds of data points) can be estimated within minutes, models with large sample size (e.g., more than thousands of data points) require excess run times, see, e.g., Aguero-Valverde and Jovanis (2010) reported that two days elapsed for their multivariate CAR model to converge after completing two chains, each with 100,000 Bayesian draws (for each parameter); and Boulieri et al. 2016 spent 20 to 27 hours to complete the 50,000 Bayesian draws (for each parameter) before reaching convergence for their Poisson Log-normal CAR model with a BYM structure. Both models were run in OpenBUGS. Run time is chiefly influenced by how fast the parameter draws converge to a stable value (if using Bayesian method) or how fast the algorithms locate the optimal solution of the likelihood function (if using maximum likelihood estimation or expected moment method). To improve computation efficiency, an analyst can consider reducing the complexity of spatial weight matrices (e.g., through sparse matrix algorithm [Finley et al. 2013]) and enhance convergence property, e.g., tweaking the acceptance rate of the M-H (so that chains converge at a faster rate) or improving parameter identification (Waller et al. 1997) by using an appropriate value for the precision parameters associated with spatial (and heterogeneity) error terms or assigning hyperpriors for these precision parameters (Eberly and Carlin 2000).

In terms of emerging opportunities, a potentially transformative one is seen in extending advanced spatial models in settings that use geo-referenced, real-time input data to make forecasts about current or near-future values (i.e., nowcasting [e.g., Lampos et al. 2015, Preis and Moat 2014]). Recent years have seen a rapid growth of real-time data with location attributes, from Google's influenza reports (which exploit Internet users' search queries), through pedestrian or cyclist route and volume data collected from smart-phone applications (Smith 2015), to vehicle and driver information streamed from connected and instrumented vehicles. Coupled with nowcasting technology, these data offer critical information for developing a real-time advisory system, such as anticipating a flu trend and offering insight for medical surveillance, or anticipating crash risk of pedestrians (or cyclists) and forewarning road users of collision risk as they navigate the network. Spatial models can enhance the regression techniques used in the nowcasting literature by controlling for spatial dependence and other interactions typically found in geo-referenced data.

Table 1. Summary Table of Spatial Models for Multivariate Count Data

Category	Author(s) (Year)	Sample Size	No. Response Levels	Model Specification	Estimation Method	How to control for correlations among multiple responses?	Software
Conditional Autoregressive Model (CAR)	Aguero-Valverde et al. (2016)	832 road segments	4	MCAR	Bayesian MCMC	MCAR structure	OpenBUGS 3.0
	Aguero-Valverde and Jovanis (2010)	7,968 segments	6	MCAR	Bayesian MCMC	MCAR structure	OpenBUGS
	Gelfand and Vounatsou (2003)	287 locations	2	MCAR	Bayesian MCMC (Gibbs sampler)	MCAR structure	-
	Jin et al. (2005)	87 counties	2	MCAR	MCMC	MCAR structure	Coded in C
	Leyland et al. (2000)	143 zip-code areas	2	Multivariate Poisson lognormal model	Iterative generalized least squares	Heterogen-eous error term	Software package MLwiN
	Song et al. (2006)	254 counties	4	MCAR	Bayesian MCMC	MCAR structure	-
	Wang and Kochelman (2013)	218 zones	2	MCAR	Bayesian MCMC	MCAR structure	WinBUGS
Multivariate Finite Mixture models	Alfo et al. (2009)	375 boroughs	2	Multivariate finite mixture models with spatial dependence	EM algorithm	Random terms generated from the convolution (BYM) model	Coded in MATLAB
	Karunanayake	150 grid cells	3	Multivariate	EM	Poisson finite	Splus/R codes

	(2007)			Poisson finite mixture models	algorithm	model structure	
Generalized Ordered-Response (GOR) Model	Castro et al. (2012)	170 intersections	1	GOR	Composite marginal likelihood	-	GAUSS
	Narayanamoorthy et al. (2013)	285 census tracts	4	GOR	Composite marginal likelihood	Multivariate normal distribution through a multinomial probit (MNP) component	GAUSS
Spatio-Temporal Models	Aldor-Noiman et al. (2012)	188 census tracts	4	Integer-valued first-order autoregressive time-series model with CAR structures	Bayesian MCMC	A Dirichlet prior placed on the rate parameters of the Poisson processes	-
	Boulieri et al. (2016)	7,932 electoral wards	2	Poisson CAR model with a BYM structure and a random walk	Bayesian MCMC	Multivariate spatially structured and unstructured effects	OpenBUGS
	Schmidt and Rodriguez (2010)	160 sites	4	Multivariate Poisson lognormal mixture model with a linear model of coregionalization (LMC)	Bayesian MCMC	Multivariate normal distribution of the error terms (which permits negative covariances)	OX

References

Aldor-Noiman, S., Brown, L.D., Fox, E.B., and Stine, R.A. (2013) Spatio-Temporal Low Count Processes with Application to Violent Crime Events. Cornell University Library. Accessed at URL: http://arxiv.org/pdf/1304.5642.pdf

Alfo, M., Nieddu, L., and Vicari, D. (2009) Finite Mixture Models for Mapping Spatially Dependent Disease Counts. *Biometrical Journal* 51(1): 84-97.

Anderson, C., Lee, D., and Dean, N. (2014) Identifying Clusters in Bayesian Disease Mapping. *Biostatsitics*.

Anselin, L. (2001) Chapter 14. Spatial Econometrics. A Companion to Theoretical Econometrics. Blackwell Publishing Ltd. http://web.pdx.edu/~crkl/WISE/SEAUG/papers/anselin01_CTE14.pdf

Anselin, L. (2010) Thirty Years of Spatial Econometrics. *Papers in Regional Science* 89, 3-25.

Atkinson, P., German, S., Sear, D. and Clark, M. (2003) Exploring the Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression. *Geographical Analysis* 35 (1): 58-82.

Aufhauser, E. and Fischer, M.M. (1985) Log-Linear Modeling and Spatial Analysis. *Environment and Planning A* 17(7), 931-951.

Bansal, P., Kockelman, K., and Wang, Y. (2015) Hybrid Electric Vehicle Ownership and Fuel Economy across Texas: Application of Spatial Models. *Transportation Research Record No.* 2495: 53-64.

Besag, J., York, J., and Molli_, A. (1991) Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* 43, 1–59.

Case, B., Clapp, J., Dubin, R., and Rodriguez, M. (2003) Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models. *The Journal of Real Estate Finance and Economics* 29 (2): 167-191.

Chakir, R., and Parent, O. (2009) Determinants of Land Use Changes: a Spatial Multinomial Probit Approach. *Papers in Regional Science* 88 (2): 327-344.

Chen, D., Wang, Y., and Kockelman, K. (2013) Where Are the Electrical Vehicles? A Spatial Model for Vehicle-Choice Count Data. *Journal of Transport Geography* 43: 181 – 188.

Chun, Y. (2008) Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems* 10 (4): 317-344.

Cressie N. A. (1991) *Statistics for Spatial Data.* John Wiley & Sons, Inc. New York.

Deutsch, C.V., Journel, A.G, (1997). *GSLIB: Geostatistical Software Library and User's Guide (Applied Geostatistics Series), Second Edition,* Oxford University Press.

Du, H. and Mulley, C. (2012) Understanding Spatial Variations in the Impact of Accessibility on Land Value Using Geographically Weighted Regression. The Journal of Transport and Land Use. Vol 5, No. 2: pp. 46 -59.

Finley, A.O., S. Banergee, and A. Gelfand (2013) spBayes for large univariate and multivariate point-referenced spatio-temporal data models. Working paper available at https://arxiv.org/pdf/1310.8192.pdf.

Gelfand, A. and Vounatsou, P. (2003) Proper Multivariate Conditional Autoregressive Models for Spatial Data Analysis. *Biostatistics* 4(1): 11-25.

Goodkind, A.L., Coggins, J.S. and Marshall, J.D. (2014) A Spatial Model of Air Pollution: The Impact of the Concentration-Response Function. Journal of the Association of Environmental and Resource Economists, Vol. 1, No. 4: pp. 451-479

Griffith, D. (2000) A Linear Regression Solution to the Spatial Autocorrelation Problem. *Journal of Geographical Systems* 2: 141-156.

Gupta, M.R. and Y. Chen (2010). *Theory and Use of the EM Algorithm*.doi:10.1561/2000000034

Hadayeghi, A., Shalaby, A., and Persaud, B. (2009) Development of Planning Level 2 Transportation Safety Tools Using Geographically Weighted Poisson Regression. *Accident Analysis & Prevention*. 42 (2):676-688.

Jin, X., Carlin, B.P., and Banerjee, S. (2005) Generalized Hierarchical Multivariate CAR Models for Areal Data. *Biometrics* 61(4): 950-961.

Krige, Danie G. (1951). "A statistical approach to some basic mine valuation problems on the Witwatersrand". *J. of the Chem., Metal. and Mining Soc. of South Africa* 52 (6): 119–139

Lampos, V., Andrew C. Miller, Steve Crossan & Christian Stefansen (2015) Advances in nowcasting influenza-like illness rates using search query logs. *Scientific Reports* 5, Article number: 12760. Available at http://www.nature.com/articles/srep12760.

LeSage, J., and Pace, K. (2009) *Introduction to Spatial Econometrics*. Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, FL.

Levine, L. (2009) Introduction to the Special Issue on Bayesian Journey to Crime Modeling. Journal of Investigative Psychology and Offender Profiling 6 (3): pp. 167-185.

Levine, N., Kim, K., and Nitz, L. (1995a) Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accident Analysis and Prevention* 27 (5): 663-674.

Levine, N., Kim, K., and Nitz, L. (1995b) Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis and Prevention* 27 (5): 675-685.

Lu, Y. (1998) Spatial Cluster Analysis for Point Data: Location Quotients versus Kernel Density, Department of Geography, State University of New York at Buffalo. http://dusk.geo.orst.edu/ucgis/web/oregon/papers/lu.htm

Miaou, S-P., Song, J., and Mallick, B. (2003) Roadway traffic crash mapping: a space-time modeling approach, *Journal of Transportation & Statistics* 6 (1): 33-58.

Nakaya, T., Fotheringham, S., Brunsdon, C. and Charlton, M. (2010) Geographically Weighted Poisson Regression for Disease Association Mapping. *Statistics in Medicine* 24 (17): 2695-2717.

Ognev-Himmelberger Y., Pearsall, H. and Rakshit, R. (2009) Concrete Evidence and Geographically Weighted Regression: a Regional Analysis of Wealth and the Land Cover in Massachusetts. *Applied Geography* 29(4): 478-487.

Páez, A. (2006) Exploring Contextual Variations in Land Use and Transport Analysis Using a 35 Probit Model with Geographical Weights. *Journal of Transport Geography* 14: 167–176.

Park, B.J. (2010) Application of Finite Mixture Models for Vehicle Crash Data Analysis. Texas A&M University Dissertation. Accessed 5/30/2016. URL: http://oaktrust.library.tamu.edu/bitstream/handle/1969.1/ETD-TAMU-2010-05-7667/PARK-DISSERTATION.pdf?sequence=2

Platt, R. (2004) Global and Local Analysis of Fragmentation in a Mountain Region of Colorado. *Agriculture, Ecosystem and Environment* 101: 207-218.

Preis, T., Helen Susannah Moat (2014) Adaptive nowcasting of influenza outbreaks using Google searches. *Royal Society Open Science* article (DOI: 10.1098/rsos.140095).

Schmidt, A.M. and Rodriguez, M.A. (2010) Modelling Multivariate Counts Varying Continuously in Space. Book chapter in *Bayesian Statistics* 9. ISBN: 9780199694587

Simoes, P. and Natario, I. (2016) Spatial Econometric Approaches for Count Data: An Overview and New Directions. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* 10 (1): 348-356.

Smith, A. (2015) Crowdsourcing Pedestrian and Cyclist Activity Data. US Department of Transportation Federal Highway Administration Report DTFHGI-11-H-00024, available at http://www.pedbikeinfo.org/cms/downloads/PBIC_WhitePaper_Crowdsourcing.pdf.

Song, J.J., Ghosh, M., Miaou, S., and Mallick, B. (2006) Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. *Journal of Multivariate Analysis* 97(1): 246- 273.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003) WinBUGS User Manual Version 1.4. URL: http://voteview.org/manual14.pdf.

Vinatier, F., Tixier, P., Duyck, P.F., and Lescourret, F. (2011) Factors and Mechanisms Explaining Spatial Heterogeneity: A Review of Methods for Insect Populations. *Methods in Ecology and Evoluation* 2 (1): 11-22.

Wang, X., and K. M. Kockelman. (2009) Application of the Dynamic Spatial Ordered Probit Model: Patterns of Land Development Change in Austin, Texas. *Papers in Regional Science* 88 (2): 345-366.

Wang, Y. and Kockelman, K (2013). A Poisson-Lognormal Conditional Autoregressive Model for Multivariate Spatial Analysis of Pedestrian Crash Counts across Neighborhoods. *Accident Analysis and Prevention* 60: 71-84.

Wang, Y., Kockelman, K., and Damien, P. (2014) A Spatial Autoregressive Multinomial Probit Model for Anticipating Land Use Change in Austin, Texas. *Annals of Regional Science* 52: 251-278.

Wang, Y., Kockelman, K., and Wang, X. (2011) Anticipating Land Use Change Using Geographically Weighted Regression Models for Discrete Response. *Transportation Research Record* No. 2245:111-123.

Weir, M., Weintraub, J., Humphreys, E., Seto, E., and Bhatia, R. (2009) An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention* 41: 137-145.

Zhao, F. and Park, N. (2004) Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. *Transportation Research Record* 1879: 99-107.