THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

WP # 0032MSS-496-2013

Date May 3, 2013

POISSON KRIGING

Victor De Oliveira
Department of Management Science and Statistics
The University of Texas at San Antonio

UTSA COLLEGE OF BUSINESS KNOWLEDGE FOR A NEW WORLD

ONE UTSA CIRCLE
SAN ANTONIO, TEXAS 78249-0631
210 458-4317 | BUSINESS.UTSA.EDU

POISSON KRIGING

Victor De Oliveira
Department of Management Science and Statistics
The University of Texas at San Antonio
San Antonio, TX 78249, USA
victor.deoliveira@utsa.edu

April 25, 2013
(preliminary version)

Abstract

This work revisits a simple model for geostatistical count data and make explicit the assumptions under which the model is constructed. We review the parameter estimators and predictors for a latent that appear in the literature, and propose new estimators and predictors. Finally, we plan to carry a detailed simulation experiment to investigate and compare the statistical properties of the different parameter estimators and predictors.

Key words: Geostatistics; Nugget effect; Ordinary kriging; Semivariogram; Spatial count data.

JEL Classifications: C21, C31, C53

1 Introduction

Spatial *count* data are routinely collected in many earth and social sciences, such as ecology, epidemiology, demography and geography. For instance, death counts due to different causes are collected on a regular basis by government agencies throughout the entire U.S. and classified according to different demographic variables, such as age, gender and race. Among the most common goals for the analysis of this kind of data are determining the effects on mortality of spatially varying risk factors (regression problems) and estimation of unobserved spatially varying quantities of interest (prediction problems). In this work I consider a model for geostatistical count data.

Early attempts to model geostatistical count data include Gotway and Stroup (1997) and McShane, Albert and Palmatier (1997), who proposed using generalized linear models and generalized estimating equations. But the statistical basis and validity of these to model geostatistical data are somewhat questionable. In addition, prediction methodology in these works is either lacking or ad-hoc, with no measures of prediction uncertainty. Many models of current use for geostatistical count data use Gaussian random fields as building blocks. The prime example is the hierarchical model proposed by Diggle, Tawn and Moyeed (1998), which can be viewed as a generalized linear mixed model. Although currently this model seems to be (arguably) the 'state-of-the-art' for modeling geostatistical count data, fitting this kind of hierarchical model is a challenging task requiring computationally intensive numerical methods, such as the EM or MCMC algorithms. This complexity is likely to preclude the use of this model by most practitioners and spatial data analysts, so it is desirable to have alternative simpler models that can be fitted using more or less traditional geostatistical methods.

We consider in this work a model for spatial count data proposed in the geostatistical literature by Monestiez, Dubroca, Bonnin, Durbec, and Guinet (2006) and Goovaerts (2005). A similar model was originally proposed by Zeger (1988) for the analysis of time series count data, and later adapted by McShane, et al. (1997) to the analysis of spatial count data. The main goal in these works was to perform regression analysis (i.e. to assess the effect of covariates on the mean response). Later, essentially the same model was introduced in the geostatistical literature by Monestiez, et al. (2006), but aimed at spatial prediction of a latent (unobserved) process; see also Goovaerts (2005), Bellier, Monestiez and Guinet (2010), and Krivoruchko, Gribov and Krause (2011) for extensions and generalizations.

In this work we revisit the aforementioned model and make explicit the assumptions under which the model is constructed. We review methods to estimate the mean function of the count data, and propose two methods to estimate the semivariogram function of the count data. We also review the method proposed by Monestiez, et al. (2006) to predict the latent process,

and proposed an alternative predictor. Finally, we plan to design a simulation experiment to investigate the properties of the different proposed estimators and predictors.

2 A Model for Spatial Count Data

Let $\{\Lambda(\mathbf{s}) : \mathbf{s} \in D\}$, with $D \subset \mathbb{R}^2$, be a *positive* random field describing the spatial variation of a quantity of interest over the domain D, usually a spatially varying intensity or risk, whose values are *not* observable. To learn about this random field, spatial information is collected on random variables $Y_1, \ldots, Y_n$ that take nonnegative integer values and whose mean values are related to $\Lambda(\cdot)$. Two examples illustrate this situation. In the Rongelap dataset analyzed by Diggle et al. (1998), $\Lambda(\mathbf{s})$ is the level of the radionuclide Caesium (^{137}Cs) at location $\mathbf{s}$, and Y_i is the number of photon emissions collected at sampling location $\mathbf{s}_i$ by a gamma-ray counter during a period of time t_i. In the farming dataset analyzed by Christensen and Waagepetersen (2002), $\Lambda(\mathbf{s})$ is the intensity of weed *(Viola arvensis)* occurrence at $\mathbf{s}$, and Y_i is the number of weeds observed within a circle of radius t_i centered at $\mathbf{s}_i$; the percentage of organic matter at each location was used as covariate information. In both examples for a set of sampling locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ within a region of interest D, a count measurement Y_i is taken, together possibly with measurements of $p \geq 1$ location-dependent covariates. The main goal in both examples is the prediction of $\Lambda(\cdot)$ throughout D based on the data $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and covariates (if available).

The proposed model for the spatial count variables $Y_1, \ldots, Y_n$ and the latent random field $\{\Lambda(\mathbf{s}) : \mathbf{s} \in D\}$ is defined hierarchically using only moments as follows:

1. For any set of distinct sampling locations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$, the counts $Y_1, \ldots, Y_n$ are conditionally independent given $\mathbf{\Lambda} = (\Lambda(\mathbf{s}_1), \ldots, \Lambda(\mathbf{s}_n))$, and

$$\mathrm{E}\{Y_i \mid \mathbf{\Lambda}\} = \mathrm{var}\{Y_i \mid \mathbf{\Lambda}\} = t_i \Lambda(\mathbf{s}_i), \qquad i = 1, \ldots, n, \tag{1}$$

   where $t_i > 0$ is known and represents the "observation/sampling effort" at sampling location $\mathbf{s}_i$. Many previous works have also assumed that, conditional on $\Lambda(\mathbf{s}_i)$, Y_i has Poisson distribution (Monestiez et al. 2006; Bellier et al. 2010), but this assumption has little bearing on the model interpretation or statistical analysis described here.

2. For any $\mathbf{s} \in D$,

$$\Lambda(\mathbf{s}) = \mu(\mathbf{s})\epsilon(\mathbf{s}),$$

   where $\mu(\mathbf{s}) > 0$ is a deterministic function representing the spatial trend in $\Lambda(\cdot)$, and $\{\epsilon(\mathbf{s}) : \mathbf{s} \in D\}$ is a positive random field with

$$\mathrm{E}\{\epsilon(\mathbf{s})\} = 1 \qquad \text{and} \qquad \mathrm{cov}\{\epsilon(\mathbf{s}), \epsilon(\mathbf{u})\} = C_\epsilon(\mathbf{s} - \mathbf{u}), \tag{2}$$

where $C_\epsilon(\mathbf{s}-\mathbf{u})$ is a *given* stationary and continuous covariance function in $\mathbb{R}^2$.

For the spatial trend we use a standard log-linear model

$$\mu(\mathbf{s}) := \exp(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s})),$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ are unknown regression parameters and $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \ldots, f_p(\mathbf{s}))'$ are known location-dependent covariates, with $f_1(\mathbf{s}) \equiv 1$ (although $\exp(\cdot)$ could be replaced by any other continuous positive function). For the covariance function of $\epsilon(\cdot)$ the common geostatistical practice is to use one of the standard positive definite functions (e.g. Monestiez et al. 2006 used the power exponential covariance function), but the question that arises is whether or not there exist a *positive* random field satisfying (2) with such function as its covariance function. This is a very difficult question with few definite answers. A way to avoid possible incompatibilities between a proposed covariance function and the required features of $\epsilon(\cdot)$ is to use a covariance function of the form (Davis et al. 1999, 2000)

$$C_\epsilon(\mathbf{s}-\mathbf{u}) := \exp(C_\delta(\mathbf{s}-\mathbf{u})) - 1, \tag{3}$$

where $C_\delta(\mathbf{s}-\mathbf{u}))$ is an arbitrary parametric covariance function in $\mathbb{R}^2$ (e.g. $C_\delta(\mathbf{s}-\mathbf{u})) = \sigma^2_\delta \exp(-||\mathbf{s}-\mathbf{u}||/\phi_\delta)$). This covariance function results when $\epsilon(\mathbf{s}) = \exp(\delta(\mathbf{s}))$ and $\delta(\cdot)$ is a Gaussian random field with mean $-C_\delta(\mathbf{0})/2$ and covariance function $C_\delta(\mathbf{s}-\mathbf{u})$. Clearly, in this case $\epsilon(\cdot)$ is positive and (2) holds. Finally the semivariogram of $\epsilon(\cdot)$ is

$$\begin{aligned} \gamma_\epsilon(\mathbf{s}-\mathbf{u}) &= \frac{1}{2}\mathrm{var}\{\epsilon(\mathbf{s}) - \epsilon(\mathbf{u})\} \\ &= \exp(C_\delta(\mathbf{0})) - \exp(C_\delta(\mathbf{s}-\mathbf{u})), \end{aligned} \tag{4}$$

when the covariance model (3) is used. We denote the model parameters by $\boldsymbol{\eta} = (\boldsymbol{\beta}', \boldsymbol{\theta}')$, where $\boldsymbol{\theta}$ are the covariance parameters that appear in $C_\epsilon(\mathbf{s}-\mathbf{u})$ and $\gamma_\epsilon(\mathbf{s}-\mathbf{u})$.

Remark 1. The above model has several attractive features. First, it allows to account for heterogeneous observation/sampling efforts which may greatly influence distributional features of the observed counts. Second, it allows the inclusion of location-dependent covariates when available. Finally, the proposed model makes no distributional assumptions since the model specification only involves moments up to second order.

The second-order structure of the latent random field $\Lambda(\cdot)$ is

$$\begin{aligned} E\{\Lambda(\mathbf{s})\} &= \mu(\mathbf{s}) \\ \mathrm{cov}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} &= \mu(\mathbf{s})\mu(\mathbf{u})C_\epsilon(\mathbf{s}-\mathbf{u}) \\ \mathrm{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} &= \frac{C_\epsilon(\mathbf{s}-\mathbf{u})}{C_\epsilon(\mathbf{0})} \\ \frac{1}{2}\mathrm{var}\{\Lambda(\mathbf{s}) - \Lambda(\mathbf{u})\} &= \mu(\mathbf{s})\mu(\mathbf{u})\gamma_\epsilon(\mathbf{s}-\mathbf{u}) + \frac{1}{2}\big(\sigma^2_\epsilon(\mu(\mathbf{s}) - \mu(\mathbf{u}))^2\big), \end{aligned}$$

so $\Lambda(\cdot)$ is a nonstationary random field when $\mu(\mathbf{s})$ is not constant, but its correlation function is always stationary. To simplify notation we write from now on μ_i instead of $\mu(\mathbf{s}_i)$.

The second-order structure of the spatial counts Y_is is obtained from (1) and the second-order structure of the random field $\Lambda(\cdot)$ by using standard mean, variance and covariance decompositions. Specifically, for any $i, j = 1, \ldots, n$

$$\begin{aligned}
\mathrm{E}\{Y_i\} &= \mathrm{E}\{\mathrm{E}(Y_i \mid \Lambda(\mathbf{s}_i))\} = t_i\mu_i, && (5)\\
\mathrm{var}\{Y_i\} &= \mathrm{E}\{\mathrm{var}(Y_i \mid \Lambda(\mathbf{s}_i))\} + \mathrm{var}\{\mathrm{E}(Y_i \mid \Lambda(\mathbf{s}_i))\} \\
&= t_i\mu_i\big(1 + \sigma_\epsilon^2 t_i\mu_i\big), && (6)\\
\mathrm{cov}\{Y_i, Y_j\} &= \mathrm{E}\{\mathrm{cov}(Y_i, Y_j \mid \Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j))\} \\
&\qquad + \mathrm{cov}\{\mathrm{E}(Y_i \mid \Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)), \mathrm{E}(Y_j \mid \Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j))\} \\
&= t_i t_j \mu_i \mu_j C_\epsilon(\mathbf{s}_i - \mathbf{s}_j), \qquad \text{for } \ i \neq j, && (7)
\end{aligned}$$

where $\sigma_\epsilon^2 := C_\epsilon(\mathbf{0})$. Note that, under this model, the second-order structure of the spatial counts Y_is is not affected by the (unspecified) bivariate distributions of $\epsilon(\cdot)$. In addition, the semivariogram of the spatial counts is give, for any $i \neq j$, by

$$\begin{aligned}
\frac{1}{2}\mathrm{var}\{Y_i - Y_j\} &= \frac{1}{2}\Big(\mathrm{var}\{Y_i\} + \mathrm{var}\{Y_j\} - 2\mathrm{cov}\{Y_i, Y_j\}\Big) \\
&= t_i t_j \mu_i \mu_j \gamma_\epsilon(\mathbf{s}_i - \mathbf{s}_j) + \frac{1}{2}\Big(\sigma_\epsilon^2\big(t_i\mu_i - t_j\mu_j\big)^2 + t_i\mu_i + t_j\mu_j\Big), && (8)
\end{aligned}$$

where (6) and (7) and were used. Several properties of this semivariogram now follow. First, the semivariogram of the spatial counts is not stationary when either the mean function is not constant or the sampling efforts are unequal. Second, the semivariogram of the spatial counts is the superposition of two positive terms, one due to the spatial variation of the latent random field $\epsilon(\cdot)$ and the other due to sampling variability of the observed counts. This in turn implies, contrary to a claim in Bellier et al. (2010), that this semivariogram is discontinuous along the diagonal line since for any $\mathbf{s}_i$

$$\lim_{\mathbf{s}_j \to \mathbf{s}_i} \frac{1}{2}\mathrm{var}\{Y_i - Y_j\} = t_i\mu_i > 0,$$

(provided $\mu(\cdot)$ is a continuous function and recalling that $\gamma_\epsilon(\cdot)$ is continuous, with $\gamma_\epsilon(\mathbf{0}) = 0$). Hence the semivariogram of the data is discontinuous along the diagonal line $\mathbf{s}_i = \mathbf{s}_j$, and the size of the discontinuity is $\mathrm{E}\{Y_i\}$. Hence, under this model there is very particular link between the mean and covariance functions of the counts.

3 Parameter Estimation

For the purposes of assessing spatial association among the spatial counts and performing prediction of the latent process $\Lambda(\cdot)$, to be described in the next section, $\mu(\mathbf{s})$ and either

$C_\epsilon(\mathbf{s}-\mathbf{u})$ or $\gamma_\epsilon(\mathbf{s}-\mathbf{u})$ need to be estimated. We consider here two methods that parallel those commonly used to estimate parameters based on continuous geostatistical data. The first method uses a two-stage approach where the regression parameters are estimated first using a pseudo-likelihood, and later the covariance parameters are estimated using a distribution-free approach based on residuals. The second approach estimates all the parameters jointly using a form of pseudo-likelihood for correlated data.

3.1 Mean Estimation

The regression parameters can be estimated by pseudo maximum likelihood, a method proposed by Gourieroux, Monfort and Trognon (1984) and used by Brännäs and Johansson (1994) and Davis et al. (1999, 2000) in the context of time series of counts. We consider below two variants based on regression models for count data (Cameron and Trivedi, 1998).

The first variant consists on using as working assumption that the count variables are independent and have Poisson distributions with means given by (5), which amounts to ignore the latent process (i.e. setting $\sigma^2_\epsilon = 0$). An estimator for $\boldsymbol{\beta}$ is then obtained as the vector $\hat{\boldsymbol{\beta}}^{\mathrm{P}} \in \mathbb{R}^p$ that maximizing the pseudo log-likelihood

$$l^{\mathrm{P}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \Big(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}_i) y_i \ - \ t_i \exp(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}_i)) \Big);$$

this estimator can be computed using the `R` function `glm`. In the context of time series of counts, Davis et al. (2000) provided a result that states conditions under which $\hat{\boldsymbol{\beta}}^{\mathrm{P}}$ is consistent and asymptotically normal, but it is unclear whether a similar result holds in the context of spatial data where infill asymptotics appear to be the most natural regime (Stein, 1999).

The second variant consists on using as working assumption that the count variables are independent and have negative binomial distributions with means and variances given by, respectively, (5) and (6), which amounts to assume that the $Y_i \mid \epsilon(\mathbf{s}_i)$ have Poisson distributions and $\epsilon(\mathbf{s}_1), \ldots, \epsilon(\mathbf{s}_n)$ are i.i.d with $\mathrm{Ga}(1/\sigma^2_\epsilon, \sigma^2_\epsilon)$ distribution. This variant also provides an estimate for the covariance parameter σ^2_ϵ. An estimator for $(\boldsymbol{\beta}, \sigma^2_\epsilon)$ is then obtained as the vector $(\hat{\boldsymbol{\beta}}^{\mathrm{NB}}, \hat{\sigma}^{2,\mathrm{NB}}_\epsilon) \in \mathbb{R}^p \times (0,\infty)$ that maximizes the pseudo log-likelihood

$$\begin{aligned} l^{\mathrm{NB}}(\boldsymbol{\beta}, \sigma^2_\epsilon) \ = \ & \sum_{i=1}^{n} \Big\{ \big(\log(\sigma^2_\epsilon) + \boldsymbol{\beta}'\mathbf{f}(\mathbf{s}_i) \big) y_i \ - \ \Big(y_i + \frac{1}{\sigma^2_\epsilon} \Big) \log\big(1 + t_i \sigma^2_\epsilon \exp(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}_i))\big) \\ & + \ \sum_{j=0}^{y_i - 1} \log\Big(\frac{1}{\sigma^2_\epsilon} + j \Big) \Big\}; \end{aligned}$$

this estimator can be computed using the `R` function `glm.nb`, which is part of the library `MASS`. Once an estimate $\hat{\boldsymbol{\beta}}$ is obtained, the resulting trend estimate is $\hat{\mu}(\mathbf{s}) = \exp(\hat{\boldsymbol{\beta}}'\mathbf{f}(\mathbf{s})), \ \mathbf{s} \in D$.

3.2 Semivariogram Estimation

The standard nonparametric geostatistical methods for semivariogram estimation, such as those described in Cressie (1993, section 2.4), are not directly applicable for the current model due to the heteroscedasticity and nonstationarity of the process of counts, and this is still so when these methods use standard residuals. Nevertheless, the basic ideas and methods used to derive these nonparametric semivariogram and covariogram estimators can be adapted and extended to the current data framework; below we describe this adaptation. The semivariogram will be estimated using the classical geostatistical approach: model-free estimates of the semivariogram are first obtained for a finite set of distances, and then the parameters of a proposed semivariogram model are estimated by least squares.

All the semivariogram and covariogram estimators to be described here require the availability of estimates for the trend at the sampling locations, to be denoted by $\hat{\mu}_i = \hat{\mu}(\mathbf{s}_i), i = 1, \ldots, n$. From these trend estimates we can compute residuals in the form of ratios, rather than the usual differences, defined as

$$R_i := \frac{Y_i}{t_i \hat{\mu}_i}, \qquad i = 1, \ldots, n.$$

Throughout this section trend estimates are treated as if they were known true values, i.e. their sampling variability is ignored, so we have that for any i

$$\mathrm{E}\{R_i\} \approx 1 \quad , \quad \mathrm{var}\{R_i\} \approx \sigma_\epsilon^2 + \frac{1}{t_i \mu_i}, \tag{9}$$

and it follows from (7) and (9) that for any $i \neq j$

$$\mathrm{cov}\{R_i, R_j\} \approx C_\epsilon(\mathbf{s}_i - \mathbf{s}_j) \quad , \quad \frac{1}{2}\mathrm{var}\{R_i - R_j\} \approx \gamma_\epsilon(\mathbf{s}_i - \mathbf{s}_j) + \frac{1}{2}\Big(\frac{t_i\mu_i + t_j\mu_j}{t_i t_j \mu_i \mu_j}\Big). \tag{10}$$

For the case when $\gamma_\epsilon(\cdot)$ is isotropic, Monestiez et al. (2006) and Bellier et al. (2010) proposed the following semivariogram estimator

$$\begin{aligned} \hat{\gamma}_\epsilon^{\mathrm{M}}(d) &= \frac{\sum_{(i,j)\in N(d)} \left(\frac{t_i t_j \hat{\mu}_i \hat{\mu}_j}{t_i\hat{\mu}_i + t_j\hat{\mu}_j}\right)\left((R_i - R_j)^2 - \frac{t_i\hat{\mu}_i + t_j\hat{\mu}_j}{t_i t_j \hat{\mu}_i \hat{\mu}_j}\right)}{2\sum_{(i,j)\in N(d)} \frac{t_i t_j \hat{\mu}_i \hat{\mu}_j}{t_i\hat{\mu}_i + t_j\hat{\mu}_j}}, \qquad d > 0 \\ &= \frac{\sum_{(i,j)\in N(d)} \left(\frac{t_i t_j \hat{\mu}_i \hat{\mu}_j}{t_i\hat{\mu}_i + t_j\hat{\mu}_j}(R_i - R_j)^2 - 1\right)}{2\sum_{(i,j)\in N(d)} \frac{t_i t_j \hat{\mu}_i \hat{\mu}_j}{t_i\hat{\mu}_i + t_j\hat{\mu}_j}}, \end{aligned} \tag{11}$$

where

$$N(d) = \{(i,j) : ||\mathbf{s}_i - \mathbf{s}_j|| \approx d\}.$$

It follows from (9) and (10) that $\hat{\gamma}_\epsilon^{\mathrm{M}}(d)$ is an (approximate) unbiased estimator of $\gamma_\epsilon(d)$. Another semivariogram estimator, also motivated by the approximate relation (10), is given by

$$\hat{\gamma}_\epsilon^{\mathrm{U}}(d) = \frac{1}{2|N(d)|} \sum_{(i,j)\in N(d)} \left((R_i - R_j)^2 - \frac{t_i\hat{\mu}_i + t_j\hat{\mu}_j}{t_i t_j \hat{\mu}_i \hat{\mu}_j}\right),$$

where $|N(d)|$ is the number of pairs in $N(d)$; $\hat{\gamma}_\epsilon^{\mathrm{U}}(d)$ is also an (approximately) unbiased estimator of $\gamma_\epsilon(d)$.

A third estimator is based on (8) and is motivated by a regression argument. From this relation and after some rearrangement of terms follow that for $i \neq j$

$$\mathrm{E}\big\{(Y_i - Y_j)^2 - (t_i\mu_i + t_j\mu_j)\big\} = 2t_it_j\mu_i\mu_j\gamma_\epsilon(d_{ij}) + (t_i\mu_i - t_j\mu_j)^2\sigma_{\epsilon 1}^2, \tag{12}$$

where $d_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$ and $\sigma_{\epsilon 1}^2 = \sigma_\epsilon^2 + 1$. If we compute the summaries

$$V_{ij} = (Y_i - Y_j)^2 - (t_i\hat{\mu}_i + t_j\hat{\mu}_j) \quad , \quad u_{ij,1} = 2t_it_j\hat{\mu}_i\hat{\mu}_j \quad , \quad u_{ij,2} = (t_i\hat{\mu}_i - t_j\hat{\mu}_j)^2,$$

then an estimate of $\gamma_\epsilon(d)$ $(d > 0)$ can be obtained by regressing V_{ij} on $(u_{ij,1}, u_{ij,2})$ without an intercept, using all the summaries for which $(i,j) \in N(d)$. The resulting estimator is

$$\hat{\gamma}_\epsilon^{\mathrm{R}}(d) = \frac{\sum_{(i,j)\in N(d)} u_{ij,2}^2 \cdot \sum_{(i,j)\in N(d)} u_{ij,1}V_{ij} \;-\; \sum_{(i,j)\in N(d)} u_{ij,1}u_{ij,2} \cdot \sum_{(i,j)\in N(d)} u_{ij,2}V_{ij}}{\sum_{(i,j)\in N(d)} u_{ij,1}^2 \cdot \sum_{(i,j)\in N(d)} u_{ij,2}^2 \;-\; \big(\sum_{(i,j)\in N(d)} u_{ij,1}u_{ij,2}\big)^2}. \tag{13}$$

This method also provides estimates for $\sigma_{\epsilon 1}^2$, one for each $d > 0$, which can be combined (say by averaging) to get a single estimate for σ_ϵ^2.

Once the model for the mean response of the counts has been estimated and its fit to the data considered adequate, (12) suggests a simple graphical way to assess the adequacy of the other model components. For a distance $0 < d < \max\{d_{ij}\}$ construct the data summaries

$$Z_{ij} = \frac{(Y_i - Y_j)^2 - (t_i\hat{\mu}_i + t_j\hat{\mu}_j)}{(t_i\hat{\mu}_i - t_j\hat{\mu}_j)^2} \quad , \quad u_{ij} = \frac{2t_it_j\hat{\mu}_i\hat{\mu}_j}{(t_i\hat{\mu}_i - t_j\hat{\mu}_j)^2}, \qquad \text{for } \; (i,j) \in N(d),$$

and draw the scatterplot of Z_{ij} versus u_{ij}. We can construct k such scatterplots corresponding to distances $d_1 < \ldots < d_k$. If the model fits the data well it is expected that all these scatterplots display close to linear relations, with slopes (approximately) increasing with distance and all having about the same intercept at the origin.

Another semivariogram estimator can be obtained by noting that the residuals follow an additive model with heteroscedastic errors. Specifically, the residuals can be written as

$$R_i = T(\mathbf{s}_i) + \xi_i,$$

where $T(\mathbf{s}) := \Lambda(\mathbf{s})/\mu(\mathbf{s})$ and $\xi_i := R_i - T(\mathbf{s}_i)$. If the Y_is and $\epsilon(\cdot)$ satisfy (1) and (2), then $T(\cdot)$ is an intrinsically stationary process with

$$\mathrm{E}\{T(\mathbf{s})\} = 1 \quad , \quad \frac{1}{2}\mathrm{var}\{T(\mathbf{s}) - T(\mathbf{u})\} = \gamma_\epsilon(\mathbf{s} - \mathbf{u}), \qquad \mathbf{s}, \mathbf{u} \in D,$$

and from (5)–(7) follow that for any i, j

$$\mathrm{E}\{\xi_i\} = 0 \quad , \quad \mathrm{cov}\{\xi_i, \xi_j\} = \begin{cases} \frac{1}{t_i \mu_i} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} ,$$

and

$$\mathrm{cov}\{\xi_i, T(\mathbf{s}_j)\} = 0.$$

Hence each residual R_i can be interpreted as the sum of an (unobserved) "signal" $T(\mathbf{s}_i)$ and a "measurement error" ξ_i, where the latter are uncorrelated among themselves and with the signal, and have unequal variances. Such a model was recently studied by Christensen (2011) who proposed methods for semivariogram estimation and signal's prediction. The semivariogram estimate proposed in that paper is

$$\hat{\gamma}_\epsilon^{\mathrm{C}}(d) = \frac{1}{2|N(d)|} \sum_{(i,j) \in N(d)} (R_i - R_j)^2 \; - \; \frac{1}{n} \sum_{i=1}^{n} \frac{1}{t_i \hat{\mu}_i}, \qquad d > 0, \tag{14}$$

which is also (approximately) unbiased.

Note that when $\mu_i = \mu$ and $t_i = t$ for all i, all of the above semivariogram estimators $\hat{\gamma}_\epsilon^{\mathrm{M}}(d), \hat{\gamma}_\epsilon^{\mathrm{U}}(d), \hat{\gamma}_\epsilon^{\mathrm{R}}(d)$ and $\hat{\gamma}_\epsilon^{\mathrm{C}}(d)$ agree and are qual to

$$\frac{1}{2|N(d)|} \sum_{(i,j) \in N(d)} (R_i - R_j)^2 \; - \; \frac{1}{t\hat{\mu}}.$$

Finally, once the model-free semivariogram estimates $\hat{\gamma}_\epsilon(d_1), \ldots, \hat{\gamma}_\epsilon(d_k)$ are obtained for distances $d_1 < \ldots < d_k$, the parameters of the proposed semivariogram model, $\gamma_\epsilon(d; \boldsymbol{\theta})$ say, are estimated by least squares:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{j=1}^{k} \big(\hat{\gamma}_\epsilon(d_j) - \gamma_\epsilon(d_j; \boldsymbol{\theta})\big)^2.$$

3.3 Estimation of σ_ϵ^2

As mentioned in Section 2, the model proposed by Monestiez et al. (2006) for spatial data is essentially the same as the model proposed by Zeger (1988) for time series data. Based on the moment relations (6) and (7), Zeger (1988) proposed estimators for σ_ϵ^2 (as well as $\rho_\epsilon(d) = C_\epsilon(d)/\sigma_\epsilon^2$) with a method-of-moments flavor. Adapting these estimators to the spatial setting result in the estimator

$$\begin{aligned} \hat{\sigma}_\epsilon^{2,\mathrm{M}} &= \frac{\sum_{i=1}^{n} \Big[(Y_i - t_i\hat{\mu}_i)^2 - t_i\hat{\mu}_i\Big]}{\sum_{i=1}^{n} t_i^2 \hat{\mu}_i^2} \\ &= \frac{\sum_{i=1}^{n} t_i^2\hat{\mu}_i^2 \Big[(R_i - 1)^2 - \frac{1}{t_i\hat{\mu}_i}\Big]}{\sum_{i=1}^{n} t_i^2 \hat{\mu}_i^2}. \end{aligned} \tag{15}$$

Also, another estimator for σ_ϵ^2 suggested by the moment relations in (9) is given by

$$\hat{\sigma}_\epsilon^{2,\mathrm{U}} = \frac{1}{n}\sum_{i=1}^{n}\Big((R_i-1)^2 - \frac{1}{t_i\hat{\mu}_i}\Big).$$

A different estimator was proposed by Brännäs and Johansson (1994) with a regression flavor, based on results from Gourieroux et al. (1984). From (6) follows that

$$(Y_i - t_i\hat{\mu}_i)^2 - t_i\hat{\mu}_i = t_i^2\hat{\mu}_i^2\sigma_\epsilon^2 + \xi_i, \qquad i = 1,\ldots,n,$$

where ξ_i are random variables with mean 0. Then the ordinary least squares estimator of σ_ϵ^2 based on the 'pseudo data' on the left hand side of the above equation is

$$\begin{aligned}\hat{\sigma}_\epsilon^{2,\mathrm{R}} &= \frac{\sum_{i=1}^n t_i^2\hat{\mu}_i^2\Big[(Y_i - t_i\hat{\mu}_i)^2 - t_i\hat{\mu}_i\Big]}{\sum_{i=1}^n t_i^4\hat{\mu}_i^4}\\ &= \frac{\sum_{i=1}^n t_i^4\hat{\mu}_i^4\Big[(R_i-1)^2 - \frac{1}{t_i\hat{\mu}_i}\Big]}{\sum_{i=1}^n t_i^4\hat{\mu}_i^4}.\end{aligned}$$

3.4 Joint Estimation: Gaussian Estimation

Gaussian estimation is a method that uses a Gaussian likelihood as a *working* objective function, for situations when the data are not Gaussian and the true likelihood is unknown. It can be viewed as a form of pseudo maximum likelihood estimation and it is potentially useful when the data are dependent. Unlike the previous methods, this method allows for the joint estimation of all model parameters. It was originally proposed by Whittle (1961) for estimation in time series, and was also used by Crowder (1985) for longitudinal binary data; see Al-Rawwash and Pourahmadi (2006) for connections between Gaussian estimation and other methods of estimation (e.g. generalized estimating equations). We plan to explore the applicability of this method for estimation in spatial models, as it does not seem to have been previously used.

Let $\mathbf{y} = (y_1,\ldots,y_n)'$ be the observed count data, $\boldsymbol{\beta}$ the regression parameters and $\boldsymbol{\theta}$ the covariance parameters that identify $C_\epsilon(\cdot)$ (such as $\boldsymbol{\theta} = (\sigma_\delta^2, \phi_\delta)$. From (5)–(7) we have

$$\mathrm{E}\{\mathbf{Y}\} = (t_1\mu(\mathbf{s}_1),\ldots,t_n\mu(\mathbf{s}_n))',$$

$$\mathrm{var}\{\mathbf{Y}\} = M\Sigma M + M,$$

where $M = M(\boldsymbol{\beta}) = \mathrm{diag}(\mathrm{E}\{\mathbf{Y}\})$ and $(\Sigma)_{ij} = (\Sigma(\boldsymbol{\theta}))_{ij} = C_\epsilon(\mathbf{s}_i - \mathbf{s}_j)$, $i,j = 1,\ldots,n$. An estimator for $(\boldsymbol{\beta},\boldsymbol{\theta})$ can then be obtained as the vector $(\hat{\boldsymbol{\beta}}^{\mathrm{G}}, \hat{\boldsymbol{\theta}}^{\mathrm{G}}) \in \mathbb{R}^p \times \mathbb{R}^2$ that minimizes the pseudo negative log-likelihood

$$\begin{aligned} l^{\mathrm{G}}(\boldsymbol{\beta},\boldsymbol{\theta}) &= \log(|\mathrm{var}\{\mathbf{Y}\}|) + (\mathbf{y} - \mathrm{E}\{\mathbf{Y}\})'(\mathrm{var}\{\mathbf{Y}\})^{-1}(\mathbf{y} - \mathrm{E}\{\mathbf{Y}\})\\ &= 2\sum_{i=1}^{n}\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}_i) + \log(|\Psi|) + (\tilde{\mathbf{r}} - \mathbf{1})'\Psi^{-1}(\tilde{\mathbf{r}} - \mathbf{1}), \end{aligned} \tag{16}$$

where $\tilde{\mathbf{r}} = \tilde{\mathbf{r}}(\boldsymbol{\beta}) = M(\boldsymbol{\beta})^{-1}\mathbf{y}$, $\mathbf{1} = (1, \ldots, 1)'$ and $\Psi = \Psi(\boldsymbol{\beta}, \boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta}) + M(\boldsymbol{\beta})^{-1}$; the second expression for the pseudo log-likelihood is obtained by writing $\text{var}\{\mathbf{Y}\} = M\Psi M$. This estimator can be computed using any algorithm for numerical maximization, e.g., the `R` function `optim`. Implementing this method is computationally intensive since requires numerical computation of $|\Psi|$ and Ψ^{-1}, so it is not practical for large datasets.

3.5 Joint Estimation: Pseudo-Composite Likelihood

A less computationally intensive variant of the above approach consists of constructing an objective function by combining bivariate normal densities, which is in the spirit of composite (or pairwise) likelihood estimation; see Varin, Reid and Firth (2011) for an overview.

Using as working assumption, for $i \neq j$, that (Y_i, Y_j) have a bivariate normal distribution with their correct second-order structure given in (5)–(7), a pseudo pairwise negative log-likelihood based on this pair of observations is

$$\begin{aligned} l_{ij}^{\text{PC}}(\boldsymbol{\beta}, \boldsymbol{\theta}) &= 2\big(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}_i) + \boldsymbol{\beta}'\mathbf{f}(\mathbf{s}_j)\big) + \log\big(\Delta_{ij}(\boldsymbol{\beta}, \boldsymbol{\theta})\big) \\ &\quad + \frac{\big(\sigma_\epsilon^2 + \frac{1}{t_j\mu_j}\big)\big(\tilde{r}_i - 1\big)^2 + \big(\sigma_\epsilon^2 + \frac{1}{t_i\mu_i}\big)\big(\tilde{r}_j - 1\big)^2 - 2C_\epsilon(\mathbf{s}_i - \mathbf{s}_j)\big(\tilde{r}_i - 1\big)\big(\tilde{r}_j - 1\big)}{\Delta_{ij}(\boldsymbol{\beta}, \boldsymbol{\theta})}, \end{aligned}$$

where $\tilde{r}_i = y_i/t_i\mu_i$ and

$$\Delta_{ij}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \big(\sigma_\epsilon^2 + \frac{1}{t_i\mu_i}\big)\big(\sigma_\epsilon^2 + \frac{1}{t_j\mu_j}\big) - C_\epsilon^2(\mathbf{s}_i - \mathbf{s}_j).$$

An estimator for $(\boldsymbol{\beta}, \boldsymbol{\theta})$ can then be obtained as the vector $(\hat{\boldsymbol{\beta}}^{\text{PC}}, \hat{\boldsymbol{\theta}}^{\text{PC}}) \in \mathbb{R}^p \times \mathbb{R}^2$ that minimizes a pseudo pairwise negative log-likelihood based on the entire data is given by

$$l^{\text{PC}}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{1 \leq i < j \leq n} w_{ij} l_{ij}^{\text{PC}}(\boldsymbol{\beta}, \boldsymbol{\theta}),$$

where w_{ij} are nonnegative weights that indicates what pairs of observations to include and seek to balance computational effort and statistical efficiency. We consider the estimators obtained by using the weights $w_{ij} = 1$ and $w_{ij} = \mathbf{1}\{||\mathbf{s}_i - \mathbf{s}_j|| < d_m\}$, with $d_m > 0$ fixed and $\mathbf{1}\{A\}$ denotes the indicator function of A.

4 Prediction of the Latent Process

We now describe a method, originally proposed by Monestiez et al. (2006) and later extended by Bellier et al. (2010), to predict the latent process $\Lambda(\cdot)$ based on the residuals $\mathbf{R} = (R_1, \ldots, R_n)'$ defined in the previous section. The method uses optimal linear prediction that mixes aspects of simple and ordinary kriging. Like simple kriging, the method assumes that both the trend

$\mu(\mathbf{s})$ and the covariance function $C_\epsilon(\mathbf{s}-\mathbf{u})$ are known, but the predictor is derived along the lines of ordinary kriging. Specifically, for any $\mathbf{s}_0 \in D$ these authors considered the family of predictors of the form

$$\mathcal{F}_0 = \Big\{ \hat{\Lambda}(\mathbf{s}_0) = \mu(\mathbf{s}_0) \cdot \mathbf{a}_0' \mathbf{R} \ : \ \mathbf{a}_0 \in \mathbb{R}^n, \ \mathbf{a}_0' \mathbf{1} = 1 \Big\}, \tag{17}$$

and we want to find the best predictor within this family. The predictors in $\mathcal{F}_0$ are unbiased for $\Lambda(\mathbf{s}_0)$ under the assumption that $\mu(\mathbf{s})$ is known (see (9)); we continue assuming this throughout the rest of this section. By a standard calculation, similar to that for deriving ordinary kriging predictors, the mean squared prediction error of a predictor in $\mathcal{F}_0$ can be written as

$$\begin{aligned}
\mathrm{MSPE}(\hat{\Lambda}(\mathbf{s}_0)) &= \mathrm{E}\{\big(\Lambda(\mathbf{s}_0) - \mu(\mathbf{s}_0) \sum_{i=1}^n a_{0i} R_i\big)^2\} \qquad (18)\\
&= \mathrm{E}\{\big(\Lambda(\mathbf{s}_0) - \mu(\mathbf{s}_0) - \mu(\mathbf{s}_0) \sum_{i=1}^n a_{0i}(R_i - 1)\big)^2\} \\
&= \mathrm{var}\{\Lambda(\mathbf{s}_0)\} + \mu^2(\mathbf{s}_0) \sum_{i=1}^n \sum_{j=1}^n a_{0i} a_{0j} \mathrm{cov}\{R_i, R_j\} - 2\mu(\mathbf{s}_0) \sum_{i=1}^n a_{0i} \mathrm{cov}\{\Lambda(\mathbf{s}_0), R_i\} \\
&= \mu^2(\mathbf{s}_0) \Big(\sigma_\epsilon^2 + \sum_{i=1}^n \frac{a_{0i}^2}{t_i \mu_i} + \sum_{i=1}^n \sum_{j=1}^n a_{0i} a_{0j} C_\epsilon(\mathbf{s}_i - \mathbf{s}_j) - 2 \sum_{i=1}^n a_{0i} C_\epsilon(\mathbf{s}_i - \mathbf{s}_0) \Big),
\end{aligned}$$

where (6) and (7) were used, as well as

$$\mathrm{cov}\{\Lambda(\mathbf{s}_0), Y_i\} = \mu(\mathbf{s}_0) t_i \mu_i C_\epsilon(\mathbf{s}_i - \mathbf{s}_0), \qquad \mathbf{s}_0 \in D; \quad i = 1, \ldots, n.$$

Then, the optimal predictor within $\mathcal{F}_0$ is obtained by minimizing over $\mathbf{a}_0 \in \mathbb{R}^n$ and $m_0 \in \mathbb{R}$ the objective function $G(\mathbf{a}_0, m_0) = \mathrm{MSPE}(\hat{\Lambda}(\mathbf{s}_0)) - 2m_0\big(\sum_{i=1}^n a_{0i} - 1\big)$, so they are obtained by solving the linear system of $n+1$ equations

$$\begin{cases} \frac{a_{0j}}{t_j \mu_j} + \sum_{i=1}^n a_{0i} C_\epsilon(\mathbf{s}_i - \mathbf{s}_j) - m_0 = C_\epsilon(\mathbf{s}_j - \mathbf{s}_0); \qquad \text{for } j = 1, \ldots, n \\ \sum_{i=1}^n a_{0i} = 1 \end{cases}, \tag{19}$$

which can be equivalently written as

$$\begin{cases} \mathbf{a}_0' \Psi - m_0 \mathbf{1}' = \mathbf{c}_0' \\ \mathbf{a}_0' \mathbf{1} = 1 \end{cases}, \tag{20}$$

with $(\Psi)_{ij} = C_\epsilon(\mathbf{s}_i - \mathbf{s}_j) + \frac{1}{t_i \mu_i} \mathbf{1}\{\mathbf{s}_i = \mathbf{s}_j\}$ and $\mathbf{c}_0' = (C_\epsilon(\mathbf{s}_1 - \mathbf{s}_0), \ldots, C_\epsilon(\mathbf{s}_n - \mathbf{s}_0))$. The solution of (20) is then unique and given by (Cressie, 1993 p. 123)

$$\mathbf{a}_0^* = \Psi^{-1}\Big(\mathbf{c}_0 + \frac{(1 - \mathbf{1}'\Psi^{-1}\mathbf{c}_0)}{\mathbf{1}'\Psi^{-1}\mathbf{1}} \mathbf{1} \Big) \qquad , \qquad m_0^* = \frac{(1 - \mathbf{1}'\Psi^{-1}\mathbf{c}_0)}{\mathbf{1}'\Psi^{-1}\mathbf{1}}.$$

The predictor $\hat{\Lambda}^*(\mathbf{s}_0) = \mu(\mathbf{s}_0)\sum_{i=1}^n a_{0i}^* R_i$ has been called the *Poisson kriging* predictor of $\Lambda(\mathbf{s}_0)$ (Monestiez et al. 2006; Goovaerts, 2006). Finally, from (18) and (19) follow that the mean squared prediction error of $\hat{\Lambda}^*(\mathbf{s}_0)$ is given by

$$\text{MSPE}(\hat{\Lambda}^*(\mathbf{s}_0)) = \mu^2(\mathbf{s}_0)\Big(\sigma_\epsilon^2 - \sum_{i=1}^n a_{0i}^* C_\epsilon(\mathbf{s}_i - \mathbf{s}_0) + m_0^*\Big).$$

Remark 2.
(a) For the case when $\mu(\mathbf{s})$ is unknown, the above expression for $\text{MSPE}(\Lambda(\mathbf{s}_0))$ is still exact when $\mu(\mathbf{s})$ is constant, while it is only approximate when $\mu(\mathbf{s})$ is nonconstant.
(b) The system of equations (20) is readily identified as the ordinary kriging system based on the covariance function $C_\epsilon(\mathbf{s}-\mathbf{u})$, for the case when the data contains measurement error with location-dependent variance of size $(t_{(\cdot)}\mu(\cdot))^{-1}$ and $\mathbf{s}_0 \neq \mathbf{s}_i$ for all i. This observation allow us to compute $\hat{\Lambda}^*(\mathbf{s}_0)$ and $\text{MSPE}(\hat{\Lambda}^*(\mathbf{s}_0))$ using some publicly available software.

4.1 An Alternative Predictor

Given that $\mu(\mathbf{s})$ is assumed known, it is also possible to compute the simple kriging predictor for $\Lambda(\cdot)$ based on the residuals $\mathbf{R}$. Specifically, for any $\mathbf{s}_0 \in D$ we consider the family of predictors of the form

$$\mathcal{G}_0 = \Big\{\hat{\Lambda}(\mathbf{s}_0) = \mu(\mathbf{s}_0)\cdot(\mathbf{b}_0'\mathbf{R} + k_0) \;:\; \mathbf{b}_0 \in \mathbb{R}^n,\; k_0 \in \mathbb{R}\Big\}, \tag{21}$$

and seek the best predictor within $\mathcal{G}_0$. By a standard calculation similar to the one detailed above (see e.g. Cressie (1993 pp 109-110)), the simple kriging predictor of $\Lambda(\mathbf{s}_0)$ based on $\mathbf{R}$ is

$$\hat{\Lambda}^\star(\mathbf{s}_0) = \mu(\mathbf{s}_0)\big(1 + \mathbf{c}_0'\Psi^{-1}(\mathbf{R} - \mathbf{1})\big),$$

and its mean squared prediction error is

$$\text{MSPE}(\hat{\Lambda}^\star(\mathbf{s}_0)) = \mu^2(\mathbf{s}_0)\big(\sigma_\epsilon^2 - \mathbf{c}_0'\Psi^{-1}\mathbf{c}_0\big).$$

We call $\hat{\Lambda}^\star(\mathbf{s}_0)$ the *simple Poisson kriging* predictor of $\Lambda(\mathbf{s}_0)$. It is clear that this predictor is unbiased for $\Lambda(\mathbf{s}_0)$, and $\text{MSPE}(\hat{\Lambda}^\star(\mathbf{s}_0)) \leq \text{MSPE}(\hat{\Lambda}^*(\mathbf{s}_0))$ since $\mathcal{F}_0 \subset \mathcal{G}_0$.

5 Summary

This work provides a detailed description of the simple model for geostatistical count data proposed by Monestiez, et al. (2006), which is more likely to be used by practitioners and spatial data analysts than the more complex hierarchical model proposed by Diggle, et al. (1998). In addition, new estimators for the mean and semivariogram functions of the count data were proposed. In follow up work we plan to investigate the statistical properties of the several estimators of model parameters and predictors of the latent process under this model.

Acknowledgements

This work was partly supported by a summer grant from the College of Business of The University of Texas at San Antonio, and by National Science Foundation grant DMS–1208896.

References

Al-Rawwash, M. and Pourahmadi, M. (2006), Gaussian Estimation and Joint Modeling of Dispersions and Correlations in Longitudinal Data. *Computer Methods and Programs in Biomedicine*, 82, 106-113.

Bellier, E., Monestiez, P. and Guinet, C. (2010), Geostatistical Modelling of Wildlife Populations: A Non-stationary Hierarchical Model for Count Data. In *geoENV VII–Geostatistics for Environmental Applications*, P.M. Atkinson and C.D. Lloyd (eds.), pp 1-12. Springer.

Brännäs, K. and Johansson, P. (1994), Time Series Count Data Regression. *Communications in Statistics–Theory and Methods*, 23, 2907-2925.

Cameron, A.C. and Trivedi, P.K. (1998), *Regression Analysis of Count Data*. Cambridge University Press.

Christensen, W.F. (2011), Filtered Kriging for Spatial Data with Heterogeneous Measurement Error Variances. *Biometrics*, 67, 947-957.

Cressie, N.A.C. (1993), *Statistics for Spatial Data* (rev. ed.). Wiley.

Crowder, M. (1985), Gaussian Estimation for Correlated Binomial Data. *Journal of the Royal Statistical Society Series B*, 47, 229-237.

Davis, R.A., Dunsmuir, W.T.M. and Wang, Y. (2000), On Autocorrelation in a Poisson Regression Model. *Biometrika*, 87, 491-505.

Davis, R.A., Wang, Y. and Dunsmuir, W.T.M. (1999), Modeling Time Series of Count Data. In *Asymptotics, Nonparametrics and Time Series*, S. Ghosh (ed.), pp 63-114. Marcel Dekker.

Diggle, P.J., Tawn, J.A. and Moyeed, R.A. (1998), Model-based Geostatistics (with discussion). *Applied Statistics*, 47, 299-326.

Goovaerts, P. (2005), Geostatistical Analysis of Disease Data: Estimation of Cancer Mortality Risk from Empirical Frequencies Using Poisson Kriging. *International Journal of Health Geographics*, 4:31.

Gourieroux, C., Monfort, A. and Trognon, A. (1984), Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica*, 52, 701-720.

Hyun, J.W., Burman, P. and Paul, D. (2012), A Regression Approach for Estimating the Parameters of the Covariance Function of a Stationary Spatial Random Process. *Journal of Statistical Planning and Inference*, 142, 2330-2344.

Krivoruchko, K., Gribov, A. and Krause, E. (2011), Multivariate Areal Interpolation for Continuous and Count Data. *Procedia Environmental Sciences*, 3, 14-19.

McShane, L.M., Albert, P.S. and Palmatier, M.A. (1997), A Latent Process Regression Model for Spatially Correlated Count Data. *Biometrics*, 53, 698-706.

Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J.-P. and Guinet, C. (2006), Geostatistical Modelling of Spatial Distribution of *Balaenoptera Physalus* in the Northwestern Mediterranean Sea from Sparse Count Data and Heterogeneous. *Ecological Modelling*, 193, 615-628.

Royle, J.A. and Wikle, C.K. (2005), Efficient Statistical Mapping of Avian Count Data. *Ecological and Environmental Statistics*, 12, 225-243.

Varin, C., Host, G. and Skare, O. (2005), Pairwise Likelihood Inference in Spatial Generalized Linear Mixed Models. *Computational Statistics and Data Analysis*, 49, 1173-1191.

Varin, C., Reid, N. and Firth, D. (2011), An Overview of Composite Likelihood Methods. *Statistica Sinica*, 21, 5-42.

Whittle, P. (1961), Gaussian Estimation in Stationary Time Series. *Bulletin of the International Statistical Institute*, 39, 1-26.

Zeger, S. (1988), A Regression Model for Time Series of Counts. *Biometrika*, 75, 621-629.