Journal of Multivariate Analysis 122 (2013) 393–408

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Hierarchical Poisson models for spatial count data

Victor De Oliveira

Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, TX 78249, USA

ARTICLE INFO

Article history:
Received 11 January 2013
Available online 23 August 2013

AMS subject classifications:
60G10
60G60
62M30

Keywords:
Copula
Fréchet–Hoeffding upper bound
Gaussian random field
Generalized linear mixed model
Geostatistics
Poisson–Gamma model
Poisson–Lognormal model

ABSTRACT

This work proposes a class of hierarchical models for geostatistical count data that includes the model proposed by Diggle et al. (1998) [13] as a particular case. For this class of models the main second-order properties of the count variables are derived, and three models within this class are studied in some detail. It is shown that for this class of models there is a close connection between the correlation structure of the counts and their overdispersions, and this property can be used to explore the flexibility of the correlation structures of these models. It is suggested that the models in this class may not be adequate to represent data consisting mostly of small counts with substantial spatial correlation. Three geostatistical count datasets are used to illustrate these issues and suggest how the results might be used to guide the selection of a model within this class.

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 US 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 models for geostatistical count data.

Unlike for *continuous* data, few models have been proposed in the literature for geostatistical count data. This is due, in part, to the scarcity of families of multivariate discrete distributions, and the fact that none of the available ones have the flexibility and mathematical tractability comparable to that of some of the families of multivariate continuous distributions (such as the Gaussian family of distributions). This scarcity is reflected by the fact that even the most influential spatial statistics textbooks either lack a treatment of models for geostatistical count data or have a very scant one, with the book by Diggle and Ribeiro [12] as a notable exception. Early works that analyze geostatistical count data include Gotway and Stroup [16] and McShane et al. [25], who proposed using generalized linear models and generalized estimating equations. But the statistical basis and validity of these approaches 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.

Most 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 et al. [13], which can be viewed as a generalized linear mixed model. This is also known as the Poisson–Lognormal model, which was initially proposed for the analysis of correlated count data by

E-mail address: victor.deoliveira@utsa.edu.

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.08.015

Aitchison and Ho [1], and used for applications in non-spatial contexts by Chan and Ledolter [4], Chib and Winkelmann [6] and Hay and Pettitt [20]. Applications of this model for the analysis of geostatistical count data appeared in Diggle et al. [13], Christensen and Waagepetersen [8], Zhang [33], Royle and Wikle [29], Christensen et al. [7], Guillot et al. [19] and Eidsvik et al. [14]. Extensions to model geostatistical zero inflated count data and multivariate count data were given, respectively, by Recta et al. [28] and Chagneau et al. [3]. Fitting this kind of hierarchical geostatistical models is a challenging task requiring numerical methods, such as EM or MCMC algorithms, and research efforts to date have almost entirely focused on fitting and computation.

As is apparent from the above, the aforementioned Poisson–Lognormal model has generated a fair amount of attention and research, and currently seems to be (arguably) the 'state-of-the-art' for modeling geostatistical count data. Nevertheless, some of the basic properties of this model are not well understood yet, and in fact the study of its main second-order properties and its adequacy to describe a variety of geostatistical count datasets have been somewhat neglected, with the work by Madsen and Dalthorp [24] as a notable exception. One salient property of this model is that the mean function of the counts may exert a substantial influence on their correlation function. This makes the model to have two features that may be undesirable for the modeling of some datasets. First, the regression parameters might be difficult to interpret and estimate because of their influence on two different aspects of the model. Second, the regression parameters induce a 'whitening effect' on the correlation structure of the counts, which in some cases renders the model unable to represent substantial spatial correlation present in some datasets. This effect is specially severe when the data consist mostly of small counts, which is precisely the case when a model accurately describing the discreteness of the data is most needed (as opposed to a model with continuous distributions intended to approximate discrete data).

In this work I propose a class of hierarchical models for geostatistical count data that includes the Poisson–Lognormal model as a particular case. The main second-order properties of the models within this class are derived, both for the latent random field in the second level of the hierarchy and for the observable counts. It is shown that for this class of models the correlation structure of the counts strongly depends on their overdispersions, which in turn are determined by the mean–variance relationship of the marginal distributions of the latent random field. An explicit expression is obtained for the correlation function of the counts in terms of their overdispersions and the correlation function of the latent random field. When there is no explicit expression for the correlation function of the latent random field, a series expansion is used to numerically investigate the second-order properties of these models. These properties should help researchers and practitioners judge the possible adequacy of any of these models to describe a particular dataset, and point to the importance of estimating overdispersion in geostatistical count data.

The above general results are used to study in some detail the second-order properties of the Poisson–Lognormal model and two Poisson–Gamma models. It is shown that one of the Poisson–Gamma models has similar second-order properties as the Poisson–Lognormal model, but the properties of the other Poisson–Gamma model are different. The findings shed light into the scope and limitations of this class of hierarchical models, and suggest that this class of models may not be adequate to describe datasets that consist mostly of small counts and display substantial spatial correlation. It is argued that these second-order properties may be used to guide the selection of a model within the proposed class and informally assess its adequacy to describe spatial count data. Some of the issues raised here are illustrated using three geostatistical count datasets that were previously analyzed in the literature.

2. A class of models for geostatistical count data

I describe here a class of models for the variation of spatial count data that generalizes the model proposed by Diggle et al. [13], and study their main second-order properties. Let $\{\Lambda(\mathbf{s}) : \mathbf{s} \in D\}$, with $D \subset \mathbb{R}^d$ and $d \geq 1$, 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 are stochastically related to $\Lambda(\cdot)$. Two examples illustrate this situation. In the Bjertorp Farm dataset analyzed by Guillot et al. [19], $\Lambda(\mathbf{s})$ represents weed intensity at location $\mathbf{s}$ and Y_i is the number of weeds observed within a rectangular frame of area t_i centered at sampling location $\mathbf{s}_i$. In the Rongelap Island dataset analyzed by Diggle et al. [13], $\Lambda(\mathbf{s})$ represents the level of radio-nuclide Caesium (^{137}Cs) at location $\mathbf{s}$ and Y_i is the number of photon emissions collected at a sampling location $\mathbf{s}_i$ by a gamma-ray counter during a period of time t_i. In these examples, for each of a set of sampling locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ within D, count measurements Y_i are taken, together possibly with measurements of 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 the covariate information, if available, but answering regression questions might be a secondary goal; see Section 5 for further details about these datasets.

The aim is for a class of models constructed to possess the following properties:

(a) The count variables Y_i have overdispersed marginal distributions, a property found to hold in many spatial (and non-spatial) count datasets.
(b) The marginal distributions of the random field $\{\Lambda(\mathbf{s}) : \mathbf{s} \in D\}$ are given by a conjectured parametric family of continuous cdfs $\mathcal{G} = \{G_{\mathbf{s}}(\cdot) : \mathbf{s} \in D\}$, where each cdf has support $[0, \infty)$.

The proposed class of models for the random vector $\mathbf{Y}$ and random field $\{\Lambda(\mathbf{s}) : \mathbf{s} \in D\}$ that satisfies the aforementioned properties is defined in terms of their family of finite-dimensional distributions. It is hierarchically specified as follows:

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

$$Y_i \mid \boldsymbol{\Lambda} \stackrel{d}{=} Y_i \mid \Lambda(\mathbf{s}_i) \sim \text{Poi}(t_i \Lambda(\mathbf{s}_i)), \quad i = 1, \ldots, n, \tag{1}$$

where $t_i > 0$ is known and represents the "sampling effort" at location $\mathbf{s}_i$.
2. For the conjectured family $\mathcal{G}$, let $T_{\mathbf{s}}(\cdot) := \Phi^{-1} \circ G_{\mathbf{s}}(\cdot)$ where $\Phi(\cdot)$ is the cdf of the standard normal distribution and $\circ$ denotes function composition. Then the random field

$$\{T_{\mathbf{s}}\big(\Lambda(\mathbf{s})\big) : \mathbf{s} \in D\}, \tag{2}$$

is assumed to be Gaussian with mean 0, variance 1 and continuous correlation function $K_\theta(\mathbf{s}, \mathbf{u})$, being also nonnegative (with θ a correlation parameter).

The random field $\Lambda(\cdot)$ may be interpreted either as a transformed Gaussian random field, similar to that studied in De Oliveira [11] (except that the transformation may depend on location), or as a random field defined using a Gaussian copula, similar to that studied in Kazianka and Pilz [22]. Non-Gaussian copulas could alternatively be used to construct models for the latent random field, but this is not considered here. Kazianka and Pilz [22] proposed the use of copulas (Gaussian and non-Gaussian) to build random field models for the response variable, rather than for the latent variable as is done here. An application of this strategy to build spatial discrete choice models was given in Sener and Bhat [30].

The family of marginal distributions of $\Lambda(\mathbf{s})$ is by construction $\mathcal{G}$, so its mean and variance functions are determined by $G_{\mathbf{s}}(\cdot)$. When available, covariate information can be incorporated into the model, say by location-dependent covariates $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \ldots, f_p(\mathbf{s}))'$, with $f_1(\mathbf{s}) \equiv 1$. In general there is little information for choosing $\mathcal{G}$, since $\Lambda(\cdot)$ is not observable, but it will be shown that this choice can be guided by the second-order properties of the observable counts.

Throughout it is assumed that the family $\mathcal{G}$ is chosen such that

$$\mu(\mathbf{s}) := E\{\Lambda(\mathbf{s})\} = \exp\big(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s})\big) \quad \text{and} \quad \text{var}\{\Lambda(\mathbf{s})\} < \infty, \quad \mathbf{s} \in D, \tag{3}$$

in which case $\text{cov}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ exists for all $\mathbf{s}, \mathbf{u} \in D$. Depending on $\mathcal{G}$ there may or may not be a closed-form expression for the covariance function of $\Lambda(\cdot)$, but the following expansion holds.

Lemma 2.1. *The covariance function of the random field $\Lambda(\cdot)$ defined in (2) can be written as*

$$\text{cov}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} = \sum_{k=1}^{\infty} a_k(\mathbf{s}) a_k(\mathbf{u}) \frac{\big(K_\theta(\mathbf{s}, \mathbf{u})\big)^k}{k!}, \quad \mathbf{s}, \mathbf{u} \in D, \tag{4}$$

with

$$a_k(\mathbf{s}) = \int_{-\infty}^{\infty} T_{\mathbf{s}}^{-1}(t) H_k(t) \phi(t) dt, \quad k = 1, 2, \ldots \tag{5}$$

where $T_{\mathbf{s}}^{-1}(\cdot) = G_{\mathbf{s}}^{-1} \circ \Phi(\cdot)$, $\phi(t)$ is the pdf of the standard normal distribution, and $H_k(t)$ is the (probabilists') Hermite polynomial of degree k ($H_1(t) = t, H_2(t) = t^2 - 1, \ldots$). Additionally, for any $\mathbf{s}, \mathbf{u} \in D$ fixed for which $K_\theta(\mathbf{s}, \mathbf{u}) \in [0, 1)$ the series (4) is absolutely convergent.

Proof. From the definition of $\Lambda(\cdot)$ in (2) it holds that

$$E\{\Lambda(\mathbf{s})\Lambda(\mathbf{u})\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_{\mathbf{s}}^{-1}(t) T_{\mathbf{u}}^{-1}(r) \phi\big(t, r; K_\theta(\mathbf{s}, \mathbf{u})\big) dt dr, \tag{6}$$

where $\phi\big(t, r; K_\theta(\mathbf{s}, \mathbf{u})\big)$ is the bivariate standard normal pdf with correlation $K_\theta(\mathbf{s}, \mathbf{u})$. Using this and the so-called Mehler's formula [17]

$$\phi\big(t, r; K_\theta(\mathbf{s}, \mathbf{u})\big) = \phi(t)\phi(r) \left(1 + \sum_{k=1}^{\infty} \frac{(K_\theta(\mathbf{s}, \mathbf{u}))^k}{k!} H_k(t) H_k(r)\right),$$

the expansion (4) follows after some manipulation. To show absolute convergence note that for any $\mathbf{s} \in D$ and $k \in \mathbb{N}$

$$|a_k(\mathbf{s})| = |E\{T_{\mathbf{s}}^{-1}(Z) H_k(Z)\}| \leq \Big(E\{(T_{\mathbf{s}}^{-1}(Z))^2\} E\{H_k^2(Z)\}\Big)^{1/2}, \quad Z \sim N(0, 1), \tag{7}$$

by the Cauchy–Schwarz inequality, and

$$E\{(T_{\mathbf{s}}^{-1}(Z))^2\} = E\{\Lambda^2(\mathbf{s})\} < \infty, \qquad E\{H_k^2(Z)\} = k!. \tag{8}$$

Hence from (7) and (8) it follows that for any $\mathbf{s}, \mathbf{u} \in D$ fixed for which $K_\theta(\mathbf{s}, \mathbf{u}) \in [0, 1)$, there exist $C > 0$ independent of k such that

$$\sum_{k=1}^{\infty}\left|a_k(\mathbf{s})a_k(\mathbf{u})\frac{\big(K_\theta(\mathbf{s}, \mathbf{u})\big)^k}{k!}\right| < C\sum_{k=1}^{\infty}\big(K_\theta(\mathbf{s}, \mathbf{u})\big)^k < \infty. \quad \square$$

Corollary 2.1. *The random field $\Lambda(\cdot)$ defined in (2) has a stationary correlation function if $K_\theta(\mathbf{s} - \mathbf{u})$ is stationary and*

$$\frac{T_{\mathbf{s}}^{-1}(t)}{\big(\mathrm{var}(\Lambda(\mathbf{s}))\big)^{1/2}} \quad \textit{does not depend on } \mathbf{s}.$$

Proof. It follows directly from (4). $\square$

Corollary 2.2 (*Conjecture*)**.** *Suppose that for every $t \in \mathbb{R}$, $G_{\mathbf{s}}(t)$ is continuous on D (as a function of $\mathbf{s}$). Then the mean and covariance functions of $\Lambda(\cdot)$ are continuous functions on D and $D \times D$, respectively.*

Proof (*Tentative*)**.** Since $\Lambda(\mathbf{s}) > 0$ for any $\mathbf{s} \in D$, it holds that

$$\begin{aligned} E\{\Lambda(\mathbf{s})\} - E\{\Lambda(\mathbf{u})\} &= \int_0^\infty \big(1 - G_{\mathbf{s}}(t)\big)dt - \int_0^\infty \big(1 - G_{\mathbf{u}}(t)\big)dt \\ &= \int_0^\infty \big(G_{\mathbf{u}}(t) - G_{\mathbf{s}}(t)\big)dt \to 0 \quad \text{as } \mathbf{s} - \mathbf{u} \to \mathbf{0}, \end{aligned}$$

because of the continuity of $G_{\mathbf{s}}(t)$, so $E\{\Lambda(\mathbf{s})\}$ is continuous on D. Now, make in (6) the change of variables

$$\begin{cases} t = w\sqrt{\dfrac{1 + K_\theta(\mathbf{s}, \mathbf{u})}{2}} + z\sqrt{\dfrac{1 - K_\theta(\mathbf{s}, \mathbf{u})}{2}} =: h_1(w, z; K_\theta(\mathbf{s}, \mathbf{u})) \\ r = w\sqrt{\dfrac{1 + K_\theta(\mathbf{s}, \mathbf{u})}{2}} - z\sqrt{\dfrac{1 - K_\theta(\mathbf{s}, \mathbf{u})}{2}} =: h_2(w, z; K_\theta(\mathbf{s}, \mathbf{u})). \end{cases}$$

Since the Jacobian of this transformation is $-\big(1 - K_\theta^2(\mathbf{s}, \mathbf{u})\big)^{1/2}$ and the image of the inverse transformation is $\mathbb{R}^2$, it follows by the change of variable formula for integrals that

$$E\{\Lambda(\mathbf{s})\Lambda(\mathbf{u})\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T_{\mathbf{s}}^{-1}\big(h_1(w, z; K_\theta(\mathbf{s}, \mathbf{u}))\big)T_{\mathbf{u}}^{-1}(h_2(w, z; K_\theta(\mathbf{s}, \mathbf{u})))\phi(w)\phi(z)dwdz,$$

which implies the integrand is continuous on $D \times D \times \mathbb{R}^2$ by the assumed continuity of $G_{\mathbf{s}}(\cdot)$, $G_{\mathbf{u}}(\cdot)$ and $K_\theta(\mathbf{s}, \mathbf{u})$. It is conjectured that the above integral converges uniformly for every $(\mathbf{s}, \mathbf{u}) \in D \times D$. This would imply that $E\{\Lambda(\mathbf{s})\Lambda(\mathbf{u})\}$ is continuous on $D \times D$, and hence so is $\mathrm{cov}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$. $\square$

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

$$E\{Y_i\} = E\{E(Y_i \mid \Lambda(\mathbf{s}_i))\} = t_i E\{\Lambda(\mathbf{s}_i)\}, \tag{9}$$

$$\begin{aligned} \mathrm{var}\{Y_i\} &= E\{\mathrm{var}(Y_i \mid \Lambda(\mathbf{s}_i))\} + \mathrm{var}\{E(Y_i \mid \Lambda(\mathbf{s}_i))\} \\ &= t_i\big(E\{\Lambda(\mathbf{s}_i)\} + t_i\mathrm{var}\{\Lambda(\mathbf{s}_i)\}\big), \end{aligned} \tag{10}$$

$$\begin{aligned} \mathrm{cov}\{Y_i, Y_j\} &= E\{\mathrm{cov}(Y_i, Y_j \mid \Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j))\} + \mathrm{cov}\{E(Y_i \mid \Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)), E(Y_j \mid \Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j))\} \\ &= t_i t_j \mathrm{cov}\{\Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)\}, \quad \text{for } i \neq j, \end{aligned} \tag{11}$$

where (1) was used in all of the above. From (9) and (10) it follows that the marginals of Y_i have overdispersed distributions, when compared to Poisson distributions, and the amount of (relative) overdispersion at location $\mathbf{s}_i$ is given by

$$\mathrm{OD}_i := \frac{\mathrm{var}\{Y_i\} - E\{Y_i\}}{E\{Y_i\}} = t_i\frac{\mathrm{var}\{\Lambda(\mathbf{s}_i)\}}{E\{\Lambda(\mathbf{s}_i)\}}. \tag{12}$$

Hence, overdispersion in the observed counts is determined by the mean–variance relation in the random field $\Lambda(\cdot)$, and is proportional to the sizes of the sampling efforts.

Depending on $\mathcal{G}$, the correlation function among the Y_i may or may not have an explicit expression and may or may not be stationary, even when the correlation function of $\Lambda(\cdot)$ is stationary. In any case the following result holds.

Proposition 2.1. *The correlation function of the count variables Y_i defined in (1) satisfies for any $i \neq j$ that*

$$\text{(a) } \operatorname{corr}\{Y_i, Y_j\} = \operatorname{corr}\{\Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)\} \times \mathrm{UB}_{ij}, \tag{13}$$

with

$$\mathrm{UB}_{ij} := \left(\left(1 + \frac{1}{\mathrm{OD}_i}\right)\left(1 + \frac{1}{\mathrm{OD}_j}\right)\right)^{-1/2}. \tag{14}$$

$$\text{(b) } 0 \leq \operatorname{corr}\{Y_i, Y_j\} < \operatorname{corr}\{\Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)\} < K_\theta(\mathbf{s}_i, \mathbf{s}_j). \tag{15}$$

Proof. (a) From (10) and (11) it holds that for any $i \neq j$

$$\begin{aligned}\operatorname{corr}\{Y_i, Y_j\} &= \frac{t_i t_j \left(\operatorname{var}\{\Lambda(\mathbf{s}_i)\}\operatorname{var}\{\Lambda(\mathbf{s}_j)\}\right)^{1/2} \operatorname{corr}\{\Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)\}}{\left([t_i(E\{\Lambda(\mathbf{s}_i)\} + t_i\operatorname{var}\{\Lambda(\mathbf{s}_i)\})][t_j(E\{\Lambda(\mathbf{s}_j)\} + t_j\operatorname{var}\{\Lambda(\mathbf{s}_j)\})]\right)^{1/2}} \\ &= \operatorname{corr}\{\Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)\} \times \mathrm{UB}_{ij}\end{aligned}$$

where UB_{ij} is given by (14).

(b) Since the joint distribution of the random variables $T_{\mathbf{s}_i}(\Lambda(\mathbf{s}_i))$ and $T_{\mathbf{s}_j}(\Lambda(\mathbf{s}_j))$ is Gaussian with nonnegative correlation, they are positively quadrant dependent.[1] This implies that $\Lambda(\mathbf{s}_i)$ and $\Lambda(\mathbf{s}_j)$ are also positively quadrant dependent, since $T_{\mathbf{s}}(\cdot)$ is an increasing function for every fixed $\mathbf{s}$, which in turn implies that $\operatorname{corr}\{\Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)\} \geq 0$ [18, Property 3]. This, (13) and the fact that $\mathrm{UB}_{ij} \in (0, 1)$ imply the left and middle inequalities in (15). The right inequality follows by a result due to Kolmogorov stating that nonlinear transformations have a 'whitening' effect on Gaussian random fields [23]. From this it follows that $\operatorname{corr}\{\Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)\} < \operatorname{corr}\{T_{\mathbf{s}_i}(\Lambda(\mathbf{s}_i)), T_{\mathbf{s}_j}(\Lambda(\mathbf{s}_j))\} = K_\theta(\mathbf{s}_i, \mathbf{s}_j)$. □

The above result (a) shows that for this class of models there is a close connection between overdispersion and correlation. The range of the correlation among the counts Y_i depends on both the range of the correlation function of $\Lambda(\cdot)$ and UB_{ij}. Moreover, $\mathrm{UB}_{ij} \to 0$ as $\mathrm{OD}_i \to 0$ or $\mathrm{OD}_j \to 0$, so low overdispersion in the count variables Y_i would restrict the possible range of their correlations, regardless of the correlation function of $\Lambda(\cdot)$.

Proposition 2.2. *Suppose that for every $t \in \mathbb{R}$, $G_{\mathbf{s}}(t)$ is continuous on D (as a function of $\mathbf{s}$), and the function $t(\mathbf{s}_i) := t_i$ is continuous on D. Then*

(a) *The covariance function of the counts Y_i is discontinuous along the 'diagonal' $\mathbf{s}_i = \mathbf{s}_j$, and the size of the discontinuity at $(\mathbf{s}_i, \mathbf{s}_i)$ is $E\{Y_i\}$.*
(b) *The size of the discontinuity of the correlation function of the counts Y_i at $(\mathbf{s}_i, \mathbf{s}_i)$ is*

$$\left(1 + \mathrm{OD}_i\right)^{-1}.$$

Proof. (a) First fix a sampling location $\mathbf{s}_i$. From the continuity of $\operatorname{cov}\{\Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)\}$ and $t(\mathbf{s}_j)$ it follows that the size of the discontinuity of $\operatorname{cov}\{Y_i, Y_j\}$ at $(\mathbf{s}_i, \mathbf{s}_i)$ is

$$\operatorname{var}\{Y_i\} - \lim_{\mathbf{s}_j - \mathbf{s}_i \to \mathbf{0}} \operatorname{cov}\{Y_i, Y_j\} = \operatorname{var}\{Y_i\} - t_i^2 \operatorname{var}\{\Lambda(\mathbf{s}_i)\} = E\{Y_i\},$$

where the first equality follows from (11) and Corollary 2.2, and the last follows from (10).

(b) For a fixed sampling location $\mathbf{s}_i$, the size of the discontinuity of $\operatorname{corr}\{Y_i, Y_j\}$ at $(\mathbf{s}_i, \mathbf{s}_i)$ is

$$\begin{aligned}1 - \lim_{\mathbf{s}_j - \mathbf{s}_i \to \mathbf{0}} \operatorname{corr}\{Y_i, Y_j\} &= 1 - \frac{\lim_{\mathbf{s}_j - \mathbf{s}_i \to \mathbf{0}} \operatorname{cov}\{Y_i, Y_j\}}{\lim_{\mathbf{s}_j - \mathbf{s}_i \to \mathbf{0}} \left(\operatorname{var}\{Y_i\}\operatorname{var}\{Y_j\}\right)^{1/2}} \\ &= 1 - \frac{t_i^2 \operatorname{var}\{\Lambda(\mathbf{s}_i)\}}{t_i\left(E\{\Lambda(\mathbf{s}_i)\} + t_i \operatorname{var}\{\Lambda(\mathbf{s}_i)\}\right)} = \left(1 + \mathrm{OD}_i\right)^{-1}. \quad \square\end{aligned}$$

The size of the discontinuity in the covariance function of the counts Y_i may be viewed as a type of location-dependent "nugget". A way to interpret this behavior is to note that the counts Y_i admit the decomposition

$$Y_i = S(\mathbf{s}_i) + \epsilon_i,$$

[1] Random variables X and Y are said to be positively quadrant dependent if

$$P(X \leq x, Y \leq y) \geq P(X \leq x)P(Y \leq y), \quad \text{for all } x, y \in \mathbb{R}.$$

.

where $S(\mathbf{s}_i) := t_i \Lambda(\mathbf{s}_i)$ and $\epsilon_i := Y_i - t_i \Lambda(\mathbf{s}_i)$. Then from (1) and (9)–(11) it follows that for every i, j

$$E\{\epsilon_i\} = 0, \qquad \operatorname{cov}\{\epsilon_i, \epsilon_j\} = \begin{cases} E\{Y_i\} & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

and

$$\operatorname{cov}\{\epsilon_i, S(\mathbf{s}_j)\} = 0.$$

Hence each datum Y_i can be interpreted as the sum of a "signal" $S(\mathbf{s}_i)$ and a "measurement error" ϵ_i, where the latter is uncorrelated with the signal and has variance equal to the signal's mean.

3. Poisson–Lognormal model

The Poisson–Lognormal model for geostatistical count data proposed by Diggle et al. [13] is a member of the family of models described in Section 2. I study here a model just slightly different from that in Diggle et al. [13], which is obtained by taking $G_{\mathbf{s}}(\cdot)$ to be the cdf of the Lognormal distribution with parameters $\nu(\mathbf{s}) = \boldsymbol{\beta}'\mathbf{f}(\mathbf{s}) - \alpha^2/2$ and $\alpha^2 > 0$ [10]. In this case

$$T_{\mathbf{s}}\big(\Lambda(\mathbf{s})\big) = \frac{\log(\Lambda(\mathbf{s})) - \nu(\mathbf{s})}{\alpha},$$

is a Gaussian random field with mean 0, variance 1 and correlation function $K_\theta(\mathbf{s}, \mathbf{u})$, say, so $\{\Lambda(\mathbf{s}) : \mathbf{s} \in D\}$ is a log-Gaussian random field whose second-order structure is given by [9]

$$\begin{aligned} E\{\Lambda(\mathbf{s})\} &= \mu(\mathbf{s}) \\ \operatorname{cov}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} &= \mu(\mathbf{s})\mu(\mathbf{u})\big(\exp(\alpha^2 K_\theta(\mathbf{s}, \mathbf{u})) - 1\big) \\ \operatorname{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} &= \frac{\exp(\alpha^2 K_\theta(\mathbf{s}, \mathbf{u})) - 1}{\exp(\alpha^2) - 1}, \quad \mathbf{s}, \mathbf{u} \in D, \end{aligned}$$

with $\mu(\mathbf{s})$ given in (3). Hence $\operatorname{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ is stationary when $K_\theta(\mathbf{s} - \mathbf{u})$ is stationary, which also follows from Corollary 2.1. The model used by Diggle et al. [13] (and many others) results when $G_{\mathbf{s}}(\cdot)$ is the cdf of the Lognormal distribution with parameters $\boldsymbol{\beta}'\mathbf{f}(\mathbf{s})$ and α^2, in which case the expressions above for $E\{\Lambda(\mathbf{s})\}$ and $\operatorname{cov}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ are multiplied by $e^{\alpha^2/2}$ and e^{α^2}, respectively. A possible issue with this parameterization is the potential identifiability problem between α^2 and the intercept β_1.

There is no closed-form expression for the marginal distribution of the counts Y_i, which might be generically called 'Poisson–Lognormal'. By the general moment relations in Section 2 it follows that

$$E\{Y_i\} = t_i \mu(\mathbf{s}_i), \qquad \operatorname{var}\{Y_i\} = E\{Y_i\}\big(1 + \sigma^2 E\{Y_i\}\big), \tag{16}$$

and

$$\mathrm{OD}_i = \sigma^2 E\{Y_i\}, \tag{17}$$

where $\sigma^2 := \exp(\alpha^2) - 1$ is a parameter that moderates overdispersion in the counts. Then for this model the amount of overdispersion in the counts Y_i is proportional to $E\{Y_i\}$, so it may vary spatially. As for the correlation function of the counts, using the second-order structure of $\Lambda(\cdot)$, (10), (11) and (13), it follows that for any $i \neq j$

$$\begin{aligned} \operatorname{corr}\{Y_i, Y_j\} &= \frac{\exp(\alpha^2 K_\theta(\mathbf{s}_i, \mathbf{s}_j)) - 1}{\left(\left(\frac{1}{E\{Y_i\}} + \exp(\alpha^2) - 1\right)\left(\frac{1}{E\{Y_j\}} + \exp(\alpha^2) - 1\right)\right)^{1/2}} \\ &= \frac{1}{\sigma^2}\Big((1 + \sigma^2)^{K_\theta(\mathbf{s}_i, \mathbf{s}_j)} - 1\Big) \times \mathrm{UB}_{ij} \\ &< \mathrm{UB}_{ij}, \end{aligned}$$

with UB_{ij} given by (14). The upper bound UB_{ij} for $\operatorname{corr}\{Y_i, Y_j\}$ is tight because $\sup\{\operatorname{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} : \mathbf{s} \neq \mathbf{u}\} = 1$. Several properties of the Poisson–Lognormal model now follow.

First, although the mean and correlation functions of $\Lambda(\cdot)$ are controlled by different parameters ($\boldsymbol{\beta}$ and (σ^2, θ), respectively), this is not the case for the count variables Y_i. The correlation among the count variables may be strongly dependent on their means when σ^2 is small, so in this case the parameters $\boldsymbol{\beta}$ affect both the means and correlations among the Y_i. For illustration I assume throughout this section that $d = 2$, $D = [0, 1] \times [0, 1]$ and $K_\theta(\mathbf{s}, \mathbf{u}) = \exp\{-\|\mathbf{s} - \mathbf{u}\|/0.3\}$, where $\|\cdot\|$ denotes Euclidean norm. Consider first the case when $E\{Y_i\}$ is constant. Fig. 1 (left) shows plots of $\operatorname{corr}\{Y_i, Y_j\}$ against $\|\mathbf{s}_i - \mathbf{s}_j\|$ for different values of $E\{Y_i\}$, when $\sigma^2 = 0.2$. In this case $\operatorname{corr}\{Y_i, Y_j\}$ displays strong sensitivity to $E\{Y_i\}$, and the size of the discontinuity at the origin is inversely proportional to $E\{Y_i\}$. Fig. 1 (right) shows a similar plot when $\sigma^2 = 4$, where in this case $\operatorname{corr}\{Y_i, Y_j\}$ is not very sensitive to $E\{Y_i\}$, and the same holds to the size of the discontinuity at the origin.

The case when $E\{Y_i\}$ is not constant is more complex to visualize because the correlation of the counts depends on distance and $(E\{Y_i\}, E\{Y_j\})$. To investigate the above issue, the sensitivity of UB_{ij} to $(E\{Y_i\}, E\{Y_j\})$ is explored since

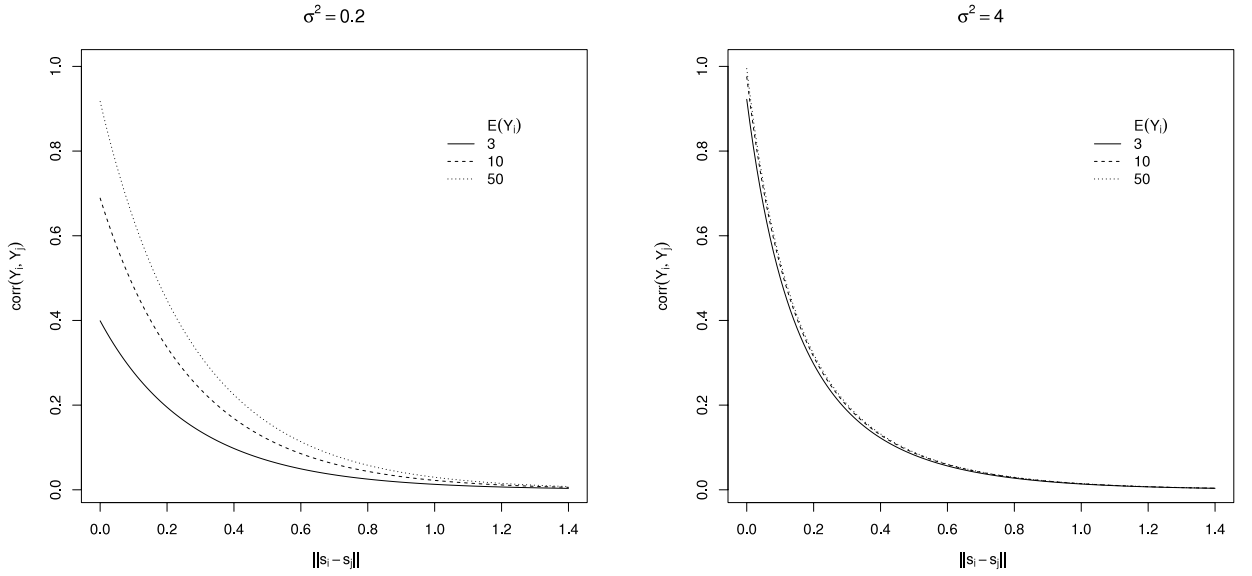

Fig. 1. Plots of correlation functions of the counts for the Poisson–Lognormal model with constant mean and $K_\theta(\mathbf{s}, \mathbf{u}) = \exp\{-\|\mathbf{s} - \mathbf{u}\|/0.3\}$, when $\sigma^2 = 0.2$ (left) and $\sigma^2 = 4$ (right).

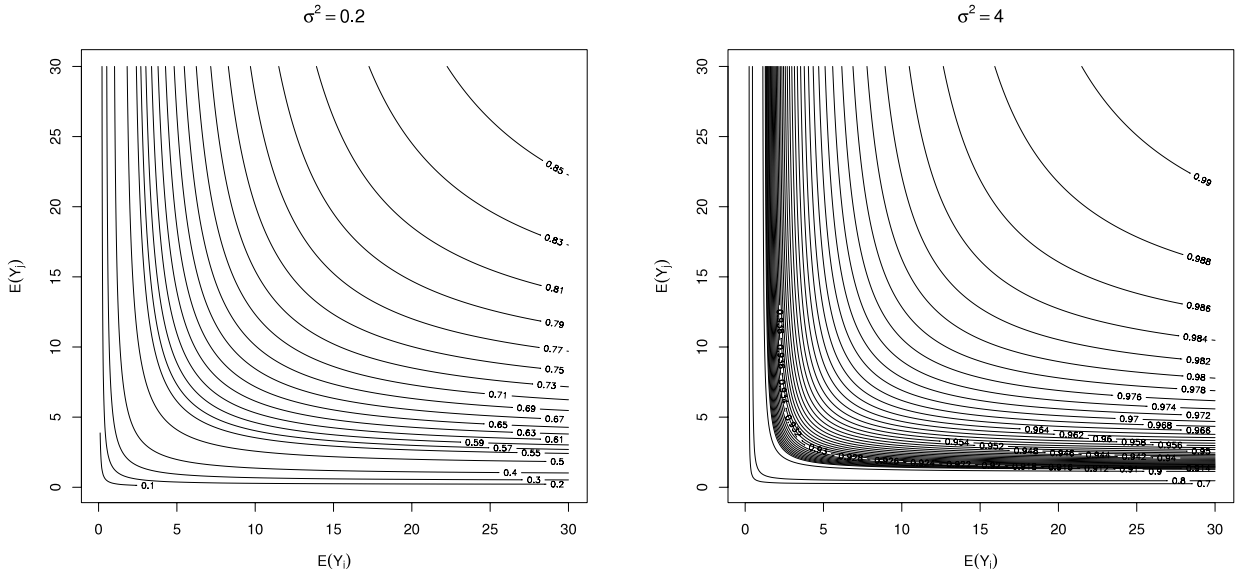

Fig. 2. Contour plots of UB_{ij} against $(E\{Y_i\}, E\{Y_j\})$ for the Poisson–Lognormal model with $K_\theta(\mathbf{s}, \mathbf{u}) = \exp\{-\|\mathbf{s} - \mathbf{u}\|/0.3\}$, when $\sigma^2 = 0.2$ (left) and $\sigma^2 = 4$ (right).

$\mathrm{corr}\{\Lambda(\mathbf{s}_i), \Lambda(\mathbf{s}_j)\}$ does not depend on $(E\{Y_i\}, E\{Y_j\})$. Fig. 2 (left) displays a contour plot of UB_{ij} against $(E\{Y_i\}, E\{Y_j\})$ when $\sigma^2 = 0.2$. In this case $\mathrm{corr}\{Y_i, Y_j\}$ is strongly sensitive to $(E\{Y_i\}, E\{Y_j\})$, unless both mean counts are large. Fig. 2 (right) displays a similar plot when $\sigma^2 = 4$, showing that in this case $\mathrm{corr}\{Y_i, Y_j\}$ is not sensitive to $(E\{Y_i\}, E\{Y_j\})$, except when both mean counts are very close to zero. The above also suggests that although $\mathrm{corr}\{Y_i, Y_j\}$ is not stationary when $E\{Y_i\}$ is not constant, the effect of $E\{Y_i\}$ on the correlation of the counts becomes small when σ^2 is moderate or large, and in this case $\mathrm{corr}\{Y_i, Y_j\}$ is close to being stationary.

Second, although the possible range of correlation for $\Lambda(\cdot)$ is $[0, 1]$, the possible range of correlation among the counts Y_i is $[0, \mathrm{UB}_{ij})$, with $\mathrm{UB}_{ij} < 1$ given by (14). When the amount of overdispersion is small everywhere (i.e. $\sigma^2 E\{Y_i\}$ is close to zero for all i), the upper bound UB_{ij} would be close to zero *regardless* of how close $\mathbf{s}_i$ and $\mathbf{s}_j$ are. In particular UB_{ij} would be close to zero when σ^2 is small and either $E\{Y_i\}$ or $E\{Y_i\}$ are close to zero; see Fig. 2 (left). This indicates that the Poisson–Lognormal model may impose severe restrictions on the correlation structure for the counts, and may be inadequate to model spatial count data consisting mostly of small counts with substantial spatial correlation among them; see Section 5. Madsen and

Dalthorp [24] discussed this issue where the upper bound for $\text{corr}\{Y_i, Y_j\}$ was also provided, although in a less interpretable form.

Finally, by noting that $\text{OD}_i = t_i\sigma^2\mu(\mathbf{s}_i)$ it follows that the behaviors discussed above could be moderated by a judicious selection of the sampling efforts t_i during data collection (if this were possible). Since σ^2 and $\mu(\mathbf{s}_i)$ are characteristics intrinsic to the random field $\Lambda(\cdot)$, the flexibility of the correlation structure of the counts under this model could be increased by choosing moderate or large sampling efforts. In addition, in this case the correlation function would be approximately stationary, which might be an appropriate or convenient feature for some datasets. A caveat though is that this potential increase in flexibility of the correlation structure of the counts needs to be balanced with the fact that large sampling efforts could render questionable an assumption in (1), namely, that the conditional distribution of Y_i depends on the latent process *only* at the sampling location $\mathbf{s}_i$.

4. Poisson–Gamma models

Hierarchical Poisson–Gamma models have been proposed for the analysis of longitudinal count data by Henderson and Shimakura [21] and Fiocco et al. [15]. In this section I describe two Poisson–Gamma spatial models with different mean–variance relationships, which in turn induce different overdispersion behaviors in the distributions of the counts.

4.1. Poisson–Gamma model 1

A new model for spatial count data is proposed here which is a member of the family of models described in Section 2. It is obtained by taking $G_{\mathbf{s}}(\cdot)$ to be the cdf of the gamma distribution with shape parameter $\alpha^2\mu(\mathbf{s})$ and scale parameter $1/\alpha^2$

$$\Lambda(\mathbf{s}) \sim \text{Ga}\left(\alpha^2\mu(\mathbf{s}), \frac{1}{\alpha^2}\right),$$

and hence $\{\Lambda(\mathbf{s}) : \mathbf{s} \in D\}$ is a kind of gamma random field with

$$E\{\Lambda(\mathbf{s})\} = \mu(\mathbf{s}) \quad \text{and} \quad \text{var}\{\Lambda(\mathbf{s})\} = \frac{\mu(\mathbf{s})}{\alpha^2}.$$

I call this the Poisson–Gamma model 1 (PG1). No closed-form expression exist for the correlation function of $\Lambda(\cdot)$ under this model, but its properties can be numerically investigated using expansion (4). From this it follows that

$$\begin{aligned}\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} &= \frac{\alpha^2}{\left(\mu(\mathbf{s})\mu(\mathbf{u})\right)^{1/2}} \sum_{k=1}^{\infty} a_k(\mathbf{s})a_k(\mathbf{u}) \frac{(K_\theta(\mathbf{s},\mathbf{u}))^k}{k!} \\ &= \frac{1}{\left(\tilde{\mu}(\mathbf{s})\tilde{\mu}(\mathbf{u})\right)^{1/2}} \sum_{k=1}^{\infty} \tilde{a}_k(\mathbf{s})\tilde{a}_k(\mathbf{u}) \frac{(K_\theta(\mathbf{s},\mathbf{u}))^k}{k!}, \end{aligned} \tag{18}$$

where $\tilde{\mu}(\mathbf{s}) := \alpha^2\mu(\mathbf{s})$ and $\tilde{a}_k(\mathbf{s}) := \alpha^2 a_k(\mathbf{s})$. The latter is given by (5) with corresponding $\text{Ga}(\alpha^2\mu(\mathbf{s}), 1)$ distribution, since $\alpha^2 G_{\mathbf{s}}^{-1}(\cdot)$ is its inverse cdf. This illustrates the fact that the correlation between random variables with marginal gamma distributions does not depend on their scale parameters, regardless of their joint distribution, so (18) may depend only on $K_\theta(\mathbf{s},\mathbf{u})$ and the shape parameters $\alpha^2\mu(\mathbf{s})$ and $\alpha^2\mu(\mathbf{u})$. Also, note that for this model the coefficient $a_k(\mathbf{s})$ in general depend on $\mathbf{s}$ when $\mu(\mathbf{s})$ is not constant, so in this case $\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ is not stationary, even when $K_\theta(\mathbf{s}-\mathbf{u})$ is.

I now numerically explore the behavior of the above correlation function, where as in Section 3 I use $D = [0,1]\times[0,1]$ and $K_\theta(\mathbf{s},\mathbf{u}) = \exp\{-\|\mathbf{s}-\mathbf{u}\|/0.3\}$. Consider first the case when $\mu(\mathbf{s}) = \mu$ (constant), so $\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ is stationary. Fig. 3 (left) shows plots of $K_\theta(\mathbf{s},\mathbf{u})$ (solid line) and $\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ (broken/dotted lines) against $\|\mathbf{s}-\mathbf{u}\|$ for different values of $\alpha^2\mu$. It shows that $\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ is not very sensitive to $\alpha^2\mu$ when the latter takes moderate or large values, and in this case $\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} \approx K_\theta(\mathbf{s},\mathbf{u})$ (an 'approximate correlation invariance', similar to that noted in [11]). The power series (18) converges very rapidly, so a good approximation results by truncating the series after a few summands (say 10) and computing the coefficients $\tilde{a}_k(\mathbf{s})$ by Gaussian quadrature. As an informal assessment of this, note that for the approximations in Fig. 3 it holds that $\sum_{k=1}^{10} \frac{\tilde{a}_k^2}{k!}$ was 0.099986, 1 and 10 for $\alpha^2\mu = 0.1$, 1 and 10, respectively (see (18)).

It follows from the definition of $\Lambda(\cdot)$ that $\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} = 0$ when $K_\theta(\mathbf{s},\mathbf{u}) = 0$, and by (15) this is its minimum value. On the other hand, from a result in [32, Theorem 2.5] and the observation after Eq. (18), it follows that the supremum of the possible correlations taken over the family of *all* bivariate distributions having $\text{Ga}(\alpha^2\mu(\mathbf{s}), 1/\alpha^2)$ and $\text{Ga}(\alpha^2\mu(\mathbf{u}), 1/\alpha^2)$ marginals, with $\mathbf{s} \neq \mathbf{u}$, is given by

$$\frac{\int_0^1 \tilde{G}_{\mathbf{s}}^{-1}(t)\tilde{G}_{\mathbf{u}}^{-1}(t)dt - \tilde{\mu}(\mathbf{s})\tilde{\mu}(\mathbf{u})}{\left(\tilde{\mu}(\mathbf{s})\tilde{\mu}(\mathbf{u})\right)^{1/2}}, \tag{19}$$

where $\tilde{G}_{\mathbf{s}}^{-1}(\cdot)$ is the inverse cdf of the $\text{Ga}\left(\tilde{\mu}(\mathbf{s}), 1\right)$ distribution. It holds that this correlation upper bound is attainable for random fields $\Lambda(\cdot)$ defined in (2) by setting $K_\theta(\mathbf{s},\mathbf{u}) = 1$ in (18) [18]. From this it follows that $\sup\{\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} : \mathbf{s} \neq$

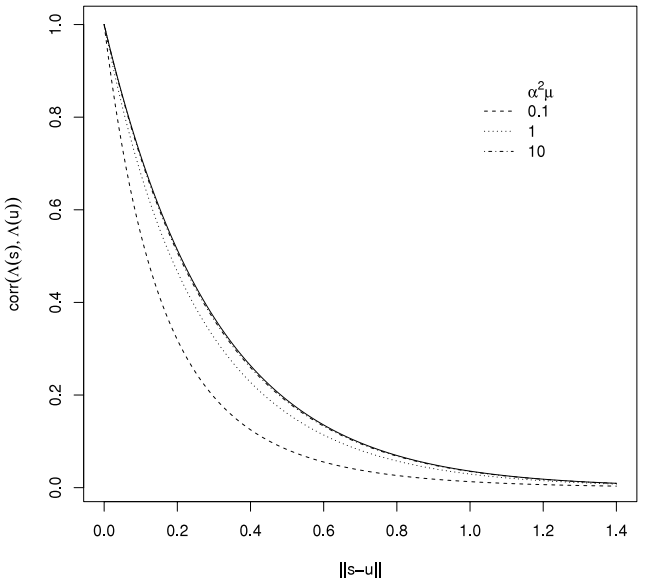

Fig. 3. Plots of correlation functions of $\Lambda(\cdot)$ for the Poisson–Gamma 1 model with constant mean and $K_\theta(\mathbf{s}, \mathbf{u}) = \exp\{-\|\mathbf{s} - \mathbf{u}\|/0.3\}$, for different shape parameters.

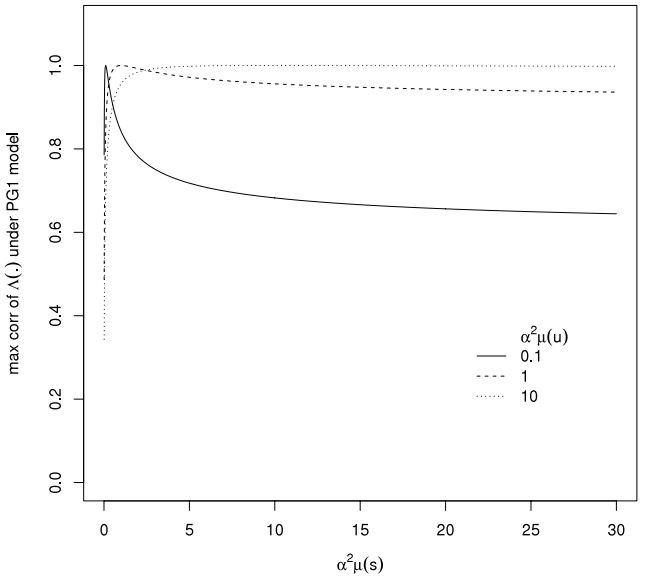

Fig. 4. Plots of maximum of the correlation functions of $\Lambda(\cdot)$ for the Poisson–Gamma 1 model as a function of one of the shape parameters, for different values of the other shape parameter.

$\mathbf{u}\} = 1$ when $\mu(\mathbf{s}) = \mu(\mathbf{u})$, since

$$\int_0^1 \big(\tilde{G}_{\mathbf{s}}^{-1}(t)\big)^2 dt = E\{\tilde{\Lambda}^2(\mathbf{s})\} = \tilde{\mu}(\mathbf{s}) + \tilde{\mu}^2(\mathbf{s}), \quad \tilde{\Lambda}(\mathbf{s}) \sim \mathrm{Ga}\big(\tilde{\mu}(\mathbf{s}), 1\big),$$

but this supremum is less than 1 when $\mu(\mathbf{s}) \neq \mu(\mathbf{u})$. To illustrate this, Fig. 4 plots (19) (computed by numerical quadrature) against $\alpha^2\mu(\mathbf{s})$ for several values of $\alpha^2\mu(\mathbf{u})$. This figure shows that a substantial restriction on the correlation range of $\Lambda(\cdot)$ might occur, but only when $\min\{\alpha^2\mu(\mathbf{s}), \alpha^2\mu(\mathbf{u})\}$ is small, say less than 1.

By a standard result on gamma mixtures of Poisson random variables, the marginal distribution of Y_i is $\mathrm{NegBin}\big(\alpha^2\mu(\mathbf{s}_i), (1 + t_i/\alpha^2)^{-1}\big)$ (called the NB1 distribution by Cameron and Trivedi [2]) with pmf

$$P\{Y_i = y\} = \frac{\Gamma\big(y + \alpha^2\mu(\mathbf{s}_i)\big)}{\Gamma\big(\alpha^2\mu(\mathbf{s}_i)\big)y!}\left(\frac{1}{1 + t_i/\alpha^2}\right)^{\alpha^2\mu(\mathbf{s}_i)}\left(1 - \frac{1}{1 + t_i/\alpha^2}\right)^{y}, \quad y = 0, 1, 2, \ldots. \tag{20}$$

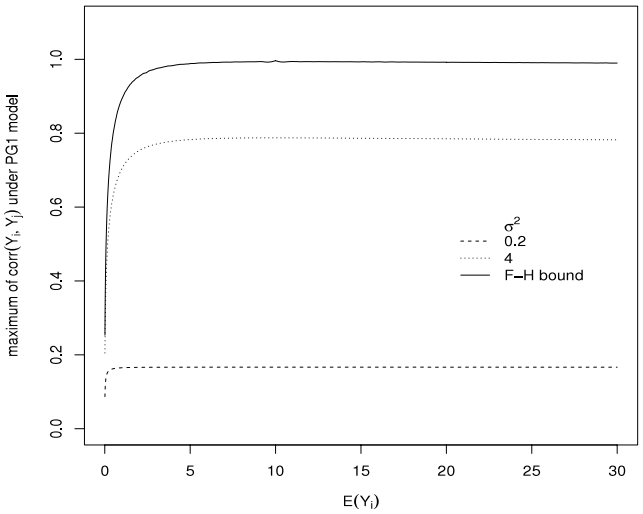

Fig. 5. Plots of correlation upper bounds for families of bivariate distributions with negative binomial marginals in (20) as a function of $E\{Y_i\}$, when $E\{Y_j\} = 10$: Fréchet–Hoeffding upper bound (solid line) and Poisson–Gamma 1 models with $\sigma^2 = 0.2$ (broken line) and $\sigma^2 = 4$ (dotted line).

Also, from the above (or the moment relations in Section 2) it follows that

$$E\{Y_i\} = t_i\mu(\mathbf{s}_i), \qquad \text{var}\{Y_i\} = E\{Y_i\}(1 + \sigma^2 t_i),$$

and

$$\text{OD}_i = \sigma^2 t_i, \tag{21}$$

with $\sigma^2 := 1/\alpha^2$. Hence, unlike the Poisson–Lognormal model, the amount of overdispersion in the count variables Y_i is independent of $\mu(\mathbf{s}_i)$ and it may vary spatially only through the sampling efforts. From this, (13) and the findings above it follows that for $i \neq j$

$$\text{corr}\{Y_i, Y_j\} = \frac{1}{\left(\tilde{\mu}(\mathbf{s}_i)\tilde{\mu}(\mathbf{s}_j)\right)^{1/2}} \sum_{k=1}^{\infty} \tilde{a}_k(\mathbf{s}_i)\tilde{a}_k(\mathbf{s}_j)\frac{(K_\theta(\mathbf{s}_i, \mathbf{s}_j))^k}{k!} \times \text{UB}_{ij}, \tag{22}$$

with UB_{ij} given by (14).

I now investigate the flexibility of the correlation structure of the counts in the PG1 model. Consider the family of *all* bivariate distributions having the same marginal negative binomial distributions as Y_i and Y_j under the PG1 model, given in (20). The maximum correlation within this family is attained by the Fréchet–Hoeffding upper bound, given by [32]

$$\frac{\int_0^1 Q_i(t)Q_j(t)dt - E\{Y_i\}E\{Y_j\}}{\left(E\{Y_i\}(1 + \sigma^2 t_i)E\{Y_j\}(1 + \sigma^2 t_j)\right)^{1/2}}, \tag{23}$$

where $Q_i(t)$ and $Q_j(t)$ are the quantile functions of Y_i and Y_j, respectively, $Q_i(t) = \inf\{x \in \mathbb{R} : P(Y_i \leq x) \geq t\}$. To explore the behavior of this bound I assume $t_i = t_j = 1$ and fix $E\{Y_j\} = 10$. Fig. 5 plots the maximum attainable correlation within this family as a function of $E\{Y_i\}$ when $\sigma^2 = 0.2$ (solid line); the corresponding graph when $\sigma^2 = 4$ is very similar (not shown). It shows that this correlation upper bound is very close to 1 almost everywhere, except when $E\{Y_i\}$ is very close to zero. Fig. 5 also plots the maximum of $\text{corr}\{Y_i, Y_j\}$ $(i \neq j)$ under the PG1 model as a function of $E\{Y_i\}$, when $\sigma^2 = 0.2$ (broken line) and $\sigma^2 = 4$ (dotted line) (both obtained by setting $K_\theta(\mathbf{s}_i, \mathbf{s}_j) = 1$ in (22)). These show that the maximum correlation of the counts under the PG1 model may be substantially smaller than what is possible for other models with the same negative binomial marginals, so severe restrictions on the attainable correlations between the counts may be imposed under this model, unless the overdispersion parameter is large.

4.2. Poisson–Gamma model 2

Here another model is proposed for spatial count data that is a member of the family of models described in Section 2, obtained also by taking $G_{\mathbf{s}}(\cdot)$ to be the cdf of a gamma distribution, but with a different parameterization. For this model it is assumed that

$$\Lambda(\mathbf{s}) \sim \text{Ga}\left(\alpha^2, \frac{\mu(\mathbf{s})}{\alpha^2}\right),$$

so $\{\Lambda(\mathbf{s}) : \mathbf{s} \in D\}$ is a kind of gamma random field with

$$E\{\Lambda(\mathbf{s})\} = \mu(\mathbf{s}) \quad \text{and} \quad \text{var}\{\Lambda(\mathbf{s})\} = \frac{\mu^2(\mathbf{s})}{\alpha^2}.$$

I call this the Poisson–Gamma model 2 (PG2). There is no closed-form expression for the correlation function of $\Lambda(\cdot)$ for this model, but the following holds.

Proposition 4.1. *For the Poisson–Gamma model 2 described above, if $K_\theta(\mathbf{s}-\mathbf{u})$ is stationary, then the correlation function of $\Lambda(\cdot)$ is also stationary and given by*

$$\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} = \alpha^2 \sum_{k=1}^{\infty} \tilde{a}_k^2 \frac{(K_\theta(\mathbf{s}-\mathbf{u}))^k}{k!}, \tag{24}$$

where the $\tilde{a}_k$ s are given by (5), but with $G_{\mathbf{s}}(\cdot)$ replaced by the cdf of the $\text{Ga}(\alpha^2, 1/\alpha^2)$ *distribution.*

Proof. The stationarity follows by noting that the correlation between random variables with marginal gamma distributions does not depend on their scale parameters, regardless of their joint distribution. Then for any $\mathbf{s} \neq \mathbf{u}$

$$\begin{aligned}\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} &= \frac{E\{\Lambda(\mathbf{s})\Lambda(\mathbf{u})\} - \mu(\mathbf{s})\mu(\mathbf{u})}{\mu(\mathbf{s})\mu(\mathbf{u})/\alpha^2} \\ &= \alpha^2\big(E\{\Upsilon(\mathbf{s})\Upsilon(\mathbf{u})\} - 1\big),\end{aligned}$$

with

$$\Upsilon(\mathbf{s}) := \frac{\Lambda(\mathbf{s})}{\mu(\mathbf{s})} \stackrel{d}{=} \frac{G_{\mathbf{s}}^{-1}}{\mu(\mathbf{s})} \circ \Phi(Z(\mathbf{s})),$$

where $G_{\mathbf{s}}^{-1}(\cdot)$ is the inverse cdf of the $\text{Ga}\big(\alpha^2, \mu(\mathbf{s})/\alpha^2\big)$ distribution, and $Z(\cdot)$ is a Gaussian random field with mean 0, variance 1 and correlation function $K_\theta(\mathbf{s}, \mathbf{u})$. Since $G_{\mathbf{s}}^{-1}(\cdot)/\mu(\mathbf{s})$ is the inverse cdf of the $\text{Ga}\big(\alpha^2, 1/\alpha^2\big)$ distribution, the random field $\Upsilon(\cdot)$ follows the same model as $\Lambda(\cdot)$, but with $\mu(\mathbf{s}) \equiv 1$. Then from (4)

$$\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} = \text{corr}\{\Upsilon(\mathbf{s}), \Upsilon(\mathbf{u})\} = \alpha^2 \sum_{k=1}^{\infty} \tilde{a}_k^2 \frac{(K_\theta(\mathbf{s}-\mathbf{u}))^k}{k!},$$

where the $\tilde{a}_k$s, being the coefficients that correspond to the random field $\Upsilon(\cdot)$ in the expansion (4), do not depend on $\mathbf{s}$ or $\mathbf{u}$. This proves the result. □

From the definition of $\Lambda(\cdot)$, $\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} = 0$ when $K_\theta(\mathbf{s}, \mathbf{u}) = 0$, and by (15) this is its minimum value. On the other hand, $\sup\{\text{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} : \mathbf{s} \neq \mathbf{u}\} = 1$, which follows from the continuity of $K_\theta(\mathbf{s}, \mathbf{u})$ and (24), since for any $\mathbf{s} \in D$

$$\sum_{k=1}^{\infty} \frac{\tilde{a}_k^2}{k!} = \text{var}\{\Upsilon(\mathbf{s})\} = \frac{1}{\alpha^2}. \tag{25}$$

Hence the correlation range of $\Lambda(\cdot)$ is $[0, 1]$.

Let $\sigma^2 := 1/\alpha^2$, which is a parameter that moderates overdispersion in the counts in the same way as the corresponding parameter of the Poisson–Lognormal (PLN) model (see below). To see how the correlation function (24) of the PG2 model compares with that of the PLN model, Fig. 6 (left) plots $\text{corr}_{\text{PG2}}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ (under PG2 model) against $\text{corr}_{\text{PLN}}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ (under PLN model) for $\sigma^2 = 0.2$ and 4, when $K_\theta(\mathbf{s}_i, \mathbf{s}_j)$ and σ^2 are the same for both models. This shows that $\text{corr}_{\text{PLN}}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} < \text{corr}_{\text{PG2}}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$, and the two are close only when the overdispersion parameter σ^2 is small. The power series (24) converges very rapidly, so a good approximation also results by truncating the series after a few summands and computing the coefficients $\tilde{a}_k$ by Gaussian quadrature. Also, identity (25) can be used to informally assess the accuracy of the approximation: for Fig. 6 it holds that $\sum_{k=1}^{10} \frac{\tilde{a}_k^2}{k!}$ was 5 and 0.2500003 when σ^2 is 0.2 and 4, respectively (recall that $\sigma^2 = 1/\alpha^2$).

Again by the standard result on gamma mixtures of Poisson random variables, the marginal distribution of Y_i is $\text{NegBin}\big(\alpha^2, (1 + t_i\mu(\mathbf{s}_i)/\alpha^2)^{-1}\big)$ (called the NB2 distribution by Cameron and Trivedi [2]) with pmf

$$P\{Y_i = y\} = \frac{\Gamma\big(y + \alpha^2\big)}{\Gamma\big(\alpha^2\big)y!}\left(\frac{1}{1 + t_i\mu(\mathbf{s}_i)/\alpha^2}\right)^{\alpha^2}\left(1 - \frac{1}{1 + t_i\mu(\mathbf{s}_i)/\alpha^2}\right)^{y}, \quad y = 0, 1, 2, \ldots \tag{26}$$

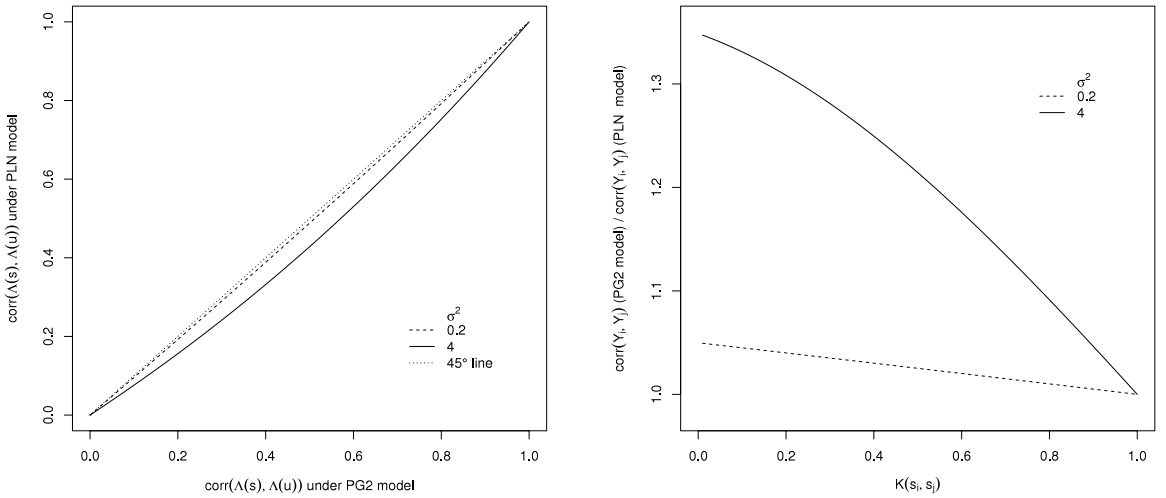

Fig. 6. Plots of corr$\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ under PG2 model against corr$\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\}$ under PLN model for two values of σ^2 (left); plots of $K_\theta(\mathbf{s}_i, \mathbf{s}_j)$ against the ratio of corr$\{Y_i, Y_j\}$ under PG2 model over corr$\{Y_i, Y_j\}$ under PLN model for two values of σ^2 (right).

and from the above (or the moment relations in Section 2) it follows that the mean, variance and overdispersion of the counts in the PG2 model are the same as those of the PLN model given in (16) and (17), provided their corresponding overdispersion parameters σ^2 match.[2] Hence UB_{ij} is also the same for both models.

For this model there is no closed-form expression for the correlation of the count variables, but using (13) and (24), it follows that for any $i \neq j$

$$\mathrm{corr}\{Y_i, Y_j\} = \frac{1}{\sigma^2}\sum_{k=1}^{\infty} \tilde{a}_k^2 \frac{(K_\theta(\mathbf{s}_i, \mathbf{s}_j))^k}{k!} \times \mathrm{UB}_{ij},$$

where UB_{ij} is given by (14). Fig. 6 (right) compares the correlation functions of the counts under the PG2 and PLN models by plotting $K_\theta(\mathbf{s}_i, \mathbf{s}_j)$ against the ratio of $\mathrm{corr}_{\mathrm{PG2}}\{Y_i, Y_j\}$ over $\mathrm{corr}_{\mathrm{PLN}}\{Y_i, Y_j\}$ for two values of the overdispersion parameter, when $K_\theta(\mathbf{s}_i, \mathbf{s}_j)$ and σ^2 are the same for both models. For small overdispersion parameters the correlation functions of the counts under both models are very close, while for large overdispersion parameters the correlation functions differ more, specially when $K_\theta(\mathbf{s}_i, \mathbf{s}_j)$ is close to zero (i.e., when $\mathbf{s}_i$ and $\mathbf{s}_j$ are far apart). All of the above shows that the correlation structures of the PG2 and PLN models have the same general behaviors and limitations, as illustrated in Figs. 1 and 2 and their explanatory text. Because of this, choosing between these models based only on second-order information does not seem feasible in practice, so other aspects of the distributions would be required for this purpose.

I now investigate further the flexibility of the correlation structure of the counts in the PG2 model, which by analogy may complement the findings for the Poisson–Lognormal model[3] given in Section 3. Consider the family of *all* bivariate distributions having the same marginal negative binomial distributions as Y_i and Y_j under the PG2 model, given in (26). The maximum correlation within this family is attained by the Fréchet–Hoeffding upper bound, which is given as before by a relation similar to (23). To explore the behavior of this bound, as before I assume $t_i = t_j = 1$, fix $E\{Y_j\} = 10$ and consider the same two values of the overdispersion parameter σ^2. Fig. 7 plots the maximum attainable correlation within this family as a function of $E\{Y_i\}$ when $\sigma^2 = 0.2$ (solid line); the corresponding graph when $\sigma^2 = 4$ is very similar (not shown). It shows that this correlation upper bound is very close to 1 almost everywhere, except when $E\{Y_i\}$ is very close to zero. Since under the PG2 (and PLN) model $\sup\{\mathrm{corr}\{\Lambda(\mathbf{s}), \Lambda(\mathbf{u})\} : \mathbf{s} \neq \mathbf{u}\} = 1$, a tight upper bound for corr$\{Y_i, Y_j\}$ under the PG2 (and PLN) model is UB_{ij} given by (14). Fig. 7 plots UB_{ij} as a function of $E\{Y_i\}$ when $\sigma^2 = 0.2$ (broken line), showing that when the overdispersion parameter is small the maximum correlation of the counts under the PG2 model is substantially smaller than what is possible for other models with the same negative binomial marginals. Fig. 7 also plots UB_{ij} as a function of $E\{Y_i\}$ when $\sigma^2 = 4$ (dotted line), showing that in this case the maximum correlation of the counts under the PG2 model is much closer to the maximum attainable correlation. Therefore, the counts under the PG2 (and PLN) model have a flexible correlation structure when the overdispersion parameter is moderate or large, but not so when that parameter is small. In the latter case the model imposes severe restrictions on the attainable correlations between the counts, regardless of how close the sampling locations are, and these restrictions are more severe when modeling datasets consisting mostly of small counts.

---

[2] For this it must hold that $\alpha^2_{\mathrm{PLN}} = \log\left(1 + \frac{1}{\alpha^2_{\mathrm{PG2}}}\right)$, where α^2_{PLN} and α^2_{PG2} denote the parameters α^2 in the PLN and PG2 models, respectively.

[3] Madsen and Dalthorp [24] did a similar comparison for the PLN model, but its validity is questionable because they assume the marginal distributions of the PLN model are negative binomial.

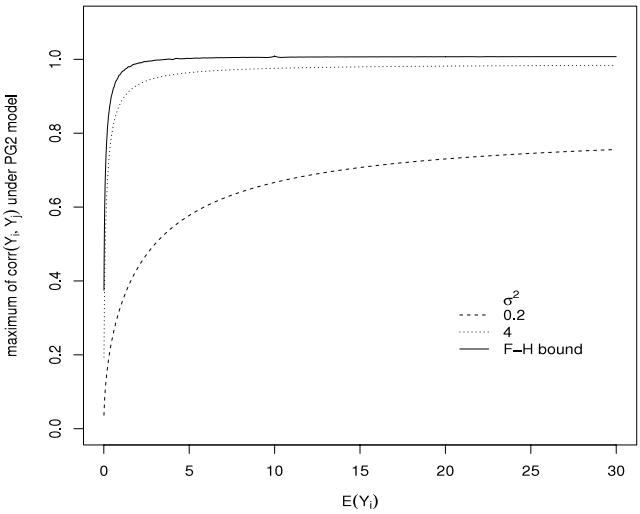

Fig. 7. Plots of correlation upper bounds as a function of $E\{Y_i\}$, when $E\{Y_j\} = 10$, for families of bivariate distributions with negative binomial marginals in (26): Fréchet–Hoeffding upper bound (solid line) and Poisson–Gamma 2 models with $\sigma^2 = 0.2$ (broken line) and $\sigma^2 = 4$ (dotted line).

5. Data illustration

A key feature of the class of hierarchical spatial models described in Section 2 is the presence of a latent random field that drives the observed count data. For some spatial datasets this latent random field has a clear physical interpretation and inference about it is the main goal. But for other datasets a clear physical interpretation for such latent random field is lacking, and its role is mostly instrumental to induce spatial association among the count data. In this section I use three spatial count datasets previously analyzed in the literature to briefly (and informally) illustrate some of the issues discussed in the previous sections.

The first example is the Bjertorp Farm dataset introduced in Section 2 which was analyzed by Guillot et al. [19] using the PLN model. It consists of weed counts measured within frames of 0.5 by 0.75 m at 100 sampling locations in an agricultural field in Sweden; the counts range from 4 to 300. In this case the sampling efforts t_i are the areas of the frames over which these counts were taken. Since these are all equal, it is assumed without loss of generality that $t_i = 1$ for all i. There are no covariates in this dataset and an exploratory analysis suggests no apparent spatial trend, so it is assumed that $E\{Y_i\}$ and $\mathrm{var}\{Y_i\}$ are both constant, which can be estimated by the sample mean and variance of the counts, 81.35 and 3678.7 respectively. The latter is a downward biased estimate though, due to dependence among the counts [27]. Fig. 8 (left) displays the empirical semivariogram of the counts Y_i. A discontinuity at the origin, a characteristic feature of semivariograms of counts in the hierarchical models of Section 2 (see Proposition 2.2), is not apparent. When fitting by least squares several isotropic semivariogram functions to this dataset (exponential and spherical), the nugget estimate was found to be zero (not shown). These make the assumption of a latent random field driving the observed counts unclear, and so is the case for any hierarchical model from Section 2 to describe the spatial variation in this dataset.

The second example is the Rongelap Island dataset introduced in Section 2, which was analyzed by Diggle et al. [13], and later by many others, using the PLN model. It consists of photon emission counts collected by a gamma-ray counter during certain periods of time at 157 sampling locations in an island that is part of the Marshal Islands; these counts range from 75 to 21390. In this case the sampling efforts t_i are the times (in seconds) over which the counts were recorded, ranging from 200 to 1800. There are no covariates in this dataset and an exploratory analysis suggests no apparent spatial trend, so it is assumed that the mean of the scaled counts Y_i/t_i is constant, which can be estimated by their sample mean, 7.6. For this dataset there is a clear physical interpretation for the latent random field (the level of radionuclide Caesium) and the connection between the observed counts and latent random field stated by the model is based on the theory of radioactive emissions (see [13] for details). Fig. 8 (right) displays the empirical semivariogram of the scaled counts Y_i/t_i for this dataset, which shows a clear discontinuity at the origin. These features suggest that using a hierarchical model from Section 2 to describe the spatial variation in this dataset is tenable. Nevertheless, ways to choose between some of the model subclasses, say between the PLN and PG1 models, are currently unknown and need to be developed; see Section 6.

The last example is a dataset on weed counts and soil magnesium amounts collected over 140 frames in an agricultural field in the Netherlands, which was discussed by Madsen and Dalthorp [24]. These authors performed the following analysis to explore the feasibility of the PLN model to describe the main second-order properties of these counts. First, a generalized linear model was fit to the weed counts using soil magnesium as covariate, finding that weed counts means ranged from 1.62 to 10.05 (counts per frame). Second, they used Taylor's power law [31], $\mathrm{var}\{Y_i\} \approx a\big(E\{Y_i\}\big)^b$, for some $a, b > 0$, to empirically

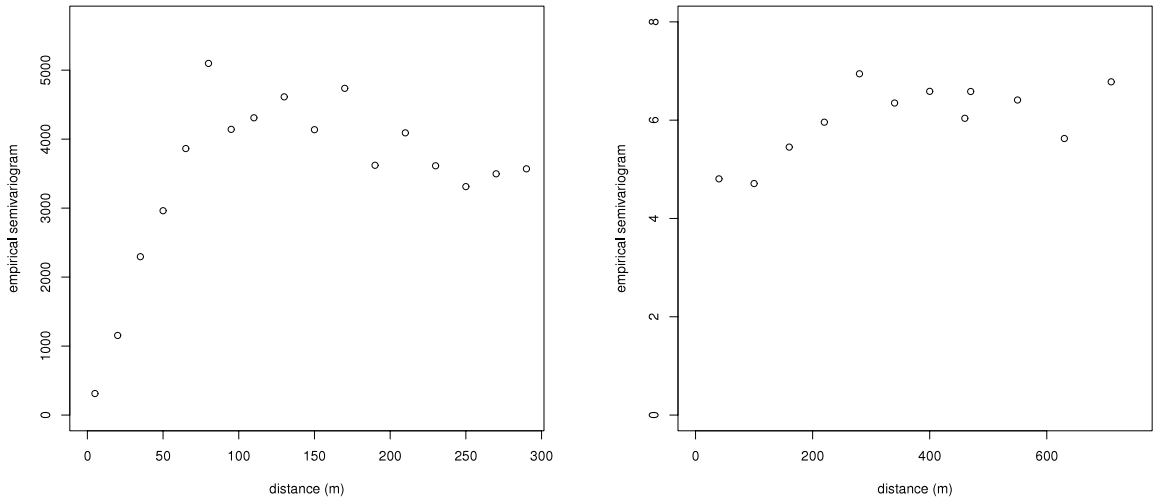

Fig. 8. Empirical semivariograms of the counts Y_i for the Bjertorp Farm dataset (left), and the scaled counts Y_i/t_i for the Rongelap Island dataset (right).

assess the mean–variance relation in the weed counts, finding that variances ranged from[4] 2.23–52.60. Finally, they fitted an isotropic spherical correlation function to the empirical correlation of the residuals, and found that these correlations were as high as 0.79. Based on these estimates the authors noted that there were 133 pairs of counts (Y_i, Y_j) for which their empirical correlations exceeded the corresponding upper bounds UB_{ij} under the PLN model[5] given in (14). This suggests the PLN model is not adequate to represent the spatial variation in this dataset. In addition, given the mostly small counts and the substantial correlation among some pairs of counts, it is likely that the PG1 and PG2 models would also be inadequate for this dataset.

6. Summary and discussion

6.1. Summary

This paper proposes a class of hierarchical models for geostatistical count data that includes the Poisson–Lognormal model as a particular case. The main second-order properties of the latent and observable variables were derived, and three members from this class of models were explored in some detail. The results of this paper shed light on the capabilities and limitations of this class of hierarchical models to describe spatial count data, which should be useful to aid researchers decide on the possible adequacy of modeling a particular dataset using a member from this class. Below I summarize the findings of the paper by listing the main second-order properties of the models in the proposed class:

1. The covariance and correlation functions of the counts depend substantially on their mean function when the overdispersion is small (Proposition 2.2; Fig. 1).
2. The correlation function of the counts is partly determined by their overdispersions, which in turn are determined by the mean–variance relationship of the marginal distributions of the latent variables (Proposition 2.1; Eq. (12)).
3. The range of the correlation function of the counts may be severely restricted, i.e., be substantially smaller than what is feasible for bivariate distributions with the same marginal distributions, when the overdispersion is small (Figs. 5 and 7).
4. A distinguishing feature of some of the models in the proposed class is their overdispersion behaviors. On the one hand the PG1 model has constant overdispersion (except possibly for varying sampling efforts), while on the other hand the PLN and PG2 models have overdispersions proportional to the mean function of the counts (Eqs. (17) and (21)).
5. The main limitation of the models in the proposed class is their apparent inability to describe datasets that consist mostly of small counts and display substantial spatial correlation.

6.2. Discussion

Below I describe some of the pending research issues needed for the practical application of the proposed class of models, which will be considered elsewhere.

Since for this class of models overdispersion of the counts was found to be determined by the mean–variance relation of the latent variables, comparison of the latter with a preliminary estimate of overdispersion in the observed counts might be

[4] It was not indicated how the law parameter a and b were obtained though.

[5] The overdispersion parameter σ^2 can be empirically estimated by noting that under the PLN model, it holds for all i that $\sigma^2 = \left(\mathrm{var}\{Y_i\} - E\{Y_i\}\right)/E^2\{Y_i\}$.

used as a guide for selecting a model within the proposed class (i.e., for the selection of $\mathcal{G}$). This in turn raises the question of how to estimate overdispersion and correlation in the observed counts in a way that is 'model-free', which appears to be a challenging task for geostatistical data that have no replicates, specially when there are covariates. The results in this article may serve as a starting point for answering this question.

An alternative approach to empirically assessing overdispersion in the observed counts, in order to guide the selection of a member within this family of models, is to work with a model in the family that allows several overdispersion behaviors. One possible such model is obtained by taking $G_{\mathbf{s}}(\cdot)$ to be the cdf of the inverse Gaussian distribution [5], with mean parameter $\mu(\mathbf{s})$ and shape parameter $\alpha^2\big(\mu(\mathbf{s})\big)^{2-\delta}$, which has the additional parameter δ. For this model it holds that $\mathrm{OD}_i = t_i\sigma^2\big(\mu(\mathbf{s}_i)\big)^{\delta}$, with $\sigma^2 := 1/\alpha^2$, so it includes several overdispersion behaviors: $\delta = 0$ results in overdispersion as in the PG1 model and $\delta = 1$ as in the PG2 model, as well as other intermediate behaviors.

It seems that goodness-of-fit methodology for geostatistical models has not been explored much, even for continuous data, to the point that it is fair to say that this area remains relatively unexplored. As a point in question it is noted that several discussants of the article by Diggle et al. [13] (Cressie and Glasbey et al.) pointed out that the PLN model does not seem to fit well the Rongelap Island dataset, and called for research on this issue. Yet, this is still due. Methodology to assess goodness-of-fit under the larger class of models proposed here is an even more relevant problem.

Finally, an important problem not considered in this paper is how to fit the models from the proposed class. Most of the currently proposed MCMC algorithms to fit the Poisson–Lognormal model may encounter serious problems for some datasets, due to very slow convergence and long running times. I expect these problems to hold, or even be compounded, for the other models in the class considered here. Hence either more efficient MCMC methods need to be designed for this class of models, such as those based on Hamiltonian dynamics [26], or other algorithms need to be adopted/developed, such as the deterministic methods based on integrated nested Laplace approximations [14]. This and several of the aforementioned important practical aspects will be reported elsewhere.

Acknowledgments

I thank two anonymous referees for helpful comments and suggestions that led to an improved article, as well as Mircea Grigoriu and Henryk Gzyl for reading a previous version of the manuscript and providing feedback. This work was partially supported by the US National Science Foundation Grants DMS–1208896 and HRD–0932339.

References

[1] J. Aitchison, C.H. Ho, The multivariate Poisson–Lognormal distribution, Biometrika 76 (1989) 643–653.
[2] A.C. Cameron, P.K. Trivedi, Regression Analysis of Count Data, Cambridge University Press, 1998.
[3] P. Chagneau, F. Mortier, N. Picard, J. Bacro, A hierarchical Bayesian model for spatial prediction of multivariate non-Gaussian random fields, Biometrics 67 (2011) 97–105.
[4] K.S. Chan, J. Ledolter, Monte Carlo EM estimation for time series models involving counts, Journal of the American Statistical Association 90 (1995) 242–252.
[5] R.S. Chhikara, J.L. Folks, The Inverse Gaussian Distribution: Theory, Methodology, and Applications, Marcel Dekker, 1989.
[6] S. Chib, R. Winkelmann, Markov Chain Monte Carlo analysis of correlated count data, Journal of Business & Economic Statistics 19 (2001) 428–435.
[7] O.F. Christensen, G.O. Roberts, M. Sköld, Robust Markov Chain Monte Carlo methods for spatial generalized linear mixed models, Journal of Computational and Graphical Statistics 15 (2006) 1–17.
[8] O.F. Christensen, R. Waagepetersen, Bayesian prediction of spatial count data using generalized linear mixed models, Biometrics 58 (2002) 280–286.
[9] E.L. Crow, K. Shimizu, Lognormal Distributions: Theory and Applications, Marcel Dekker, 1988.
[10] R.A. Davis, W.T.M. Dunsmuir, Y. Wang, On autocorrelation in a Poisson regression model, Biometrika 87 (2000) 491–505.
[11] V. De Oliveira, A note on the correlation structure of transformed Gaussian random fields, Australian and New Zealand Journal of Statistics 45 (2003) 353–366.
[12] P.J. Diggle, P.J. Ribeiro, Model-Based Geostatistics, Springer-Verlag, 2007.
[13] P.J. Diggle, J.A. Tawn, R.A. Moyeed, Model-based geostatistics (with discussion), Journal of the Royal Statistical Society: Series C 47 (1998) 299–326.
[14] J. Eidsvik, S. Martino, H. Rue, Approximate Bayesian inference in spatial generalized linear mixed models, Scandinavian Journal of Statistics 36 (2009) 1–22.
[15] M. Fiocco, H. Putter, J.C. Van Houwelingen, A new serially correlated gamma-frailty process for longitudinal count data, Biostatistics 10 (2009) 245–257.
[16] C.A. Gotway, W.W. Stroup, A generalized linear model approach to spatial data analysis and prediction, Journal of Agricultural, Biological, and Environmental Statistics 2 (1997) 157–178.
[17] C.W.J. Granger, P. Newbold, Forecasting transformed series, Journal of the Royal Statistical Society: Series B 38 (1976) 189–203.
[18] M. Grigoriu, Multivariate distributions with specified marginals: applications to wind engineering, Journal of Engineering Mechanics 133 (2007) 174–184.
[19] G. Guillot, N. Lorén, M. Rudemo, Spatial prediction of weed intensities from exact count data and image-based estimates, Journal of the Royal Statistical Society: Series C 58 (2009) 525–542.
[20] J.L. Hay, A.N. Pettitt, Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease, Biostatistics 2 (2001) 433–444.
[21] R. Henderson, S. Shimakura, A serially correlated gamma frailty model for longitudinal count data, Biometrika 90 (2003) 355–366.
[22] H. Kazianka, J. Pilz, Copula-based geostatistical modeling of continuous and discrete data including covariates, Stochastic Environmental Research and Risk Assessment 24 (2010) 661–673.
[23] R.A. Koyak, On measuring internal dependence in a set of random variables, The Annals of Statistics 15 (1987) 1215–1228.
[24] L. Madsen, D. Dalthorp, Simulating correlated count data, Environmental and Ecological Statistics 14 (2007) 129–148.
[25] L.M. McShane, P.S. Albert, M.A. Palmatier, A latent process regression model for spatially correlated count data, Biometrics 53 (1997) 698–706.
[26] R.M. Neal, MCMC using Hamiltonian dynamics, in: S. Brooks, A. Gelman, G.L. Jones, X.-L. Meng (Eds.), Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC, 2011, pp. 113–162.
[27] D.B. Percival, Three curious properties of the sample variance and autocovariance for stationary processes with unknown mean, The American Statistician 47 (1993) 274–276.

[28] V. Recta, M. Haran, J.L. Rosenberger, A two-stage model for incidence and prevalence in point-level spatial count data, Environmetrics 23 (2012) 162–174.
[29] J.A. Royle, C.K. Wikle, Efficient statistical mapping of avian count data, Environmental and Ecological Statistics 12 (2005) 225–243.
[30] I.N. Sener, C.R. Bhat, Flexible spatial dependence structures for unordered multinomial choice models: formulation and application to Teenagers' activity participation, Transportation 39 (2012) 657–683.
[31] L.R. Taylor, Aggregation, variance and the mean, Nature 189 (1961) 732–737.
[32] W. Whitt, Bivariate distributions with given marginals, The Annals of Statistics 4 (1976) 1280–1289.
[33] H. Zhang, On estimation and prediction for spatial generalized linear mixed models, Biometrics 58 (2002) 129–136.