SORT 45 (1) January-June 2021, 67-90 DOI: 10.2436/20.8080.02.110

The unilateral spatial autoregressive process for the regular lattice two-dimensional spatial discrete data

Azmi Chutoo[1], Dimitris Karlis[2], Naushad Mamode Khan[1] and Vandna Jowaheer[3]

Abstract

This paper proposes a generalized framework to analyze spatial count data under a unilateral regular lattice structure based on thinning type models. We start from the simple spatial integer-valued auto-regressive model of order 1. We extend this model in certain directions. First, we consider various distributions as choices for the innovation distribution to allow for additional overdispersion. Second, we allow for use of covariate information, leading to a non-stationary model. Finally, we derive and use other models related to this simple one by considering simplification on the existing model. Inference is based on conditional maximum likelihood approach. We provide simulation results under different scenarios to understand the behaviour of the conditional maximum likelihood. A real data application is also provided. Remarks on how the results extend to other families of models are also given.

MSC: 62H11, 62M30.

Keywords: Unilateral, Spatial, Regular, Lattice, Thinning.

1 Introduction

Problems with spatial count data occur in several disciplines. For example, consider the Human West Nile virus counts spread (Tevie, Bohara and Valdez, 2014), the infant death syndrome for the counties in North Carolina (Cressie and Chan, 1989), the number of vehicle burglary incidents in counties of Texas (Chun, 2014) and most recently, the cases and/or deaths of COVID-19 outbreak.

Spatial data usually viewed as an aggregation or average of the events of interest emanates from a lattice structure. There are two main broad ways of representing the spatial observations (Cressie, 1993). The first, and most common way, is when the ob-

[1] Department of Economics and Statistics, University of Mauritius, Mauritius.

[2] Department of Statistics, Athens University of Economics and Business, Greece.

[3] Department of Mathematics, University of Mauritius, Mauritius.

Received: December 2020
Accepted: April 2021

servation is in the form of a single indexed variable obtained from an areal unit k within some defined boundaries and in which case the spatial event is denoted as Y_k, for the kth areal unit. In the second way, the spatial count observation is indexed in two dimensional forms, in terms of the location or site coordinate (or also termed as the latitude and longitude position (i, j)), denoted as Y_{ij}, situated over a regular or irregular grid. Under both representations, the spatial count observation is supplemented by a neighbourhood structure, defined by terms of areal units or sites within the lattice structure. The second form of representation is useful since the two-dimensional representation considers all border cells in the region of interest (see, e.g. Tjøstheim, 1978a; Basu and Reinsel, 1993). Details on these representations can be found in Cressie (1993, Chap 6).

In the analysis of spatial data, it is important to investigate the spatial dependence between observations from the different neighbouring areal units or sites. Next, such analysis can also shed light on the possible factors or effects influencing the spatial observations and these can include variables such as the distance metric, elevation, slope, rock type and land use fault types (see, e.g. Tobler, 1969).

Several models have been studied in the literature to analyze spatial processes. The majority of the literature treats spatial continuous or discrete processes involving areal units, while only few papers consider the spatial data with a coordinate system, especially for the discrete case. Besides, such spatial observations are seen mostly in the agricultural, disease mapping, environmental and in the field of criminology. Specifically, in agriculture, the plantation field is usually split into small areas or say, square grids or cells with location (i, j) and wherein each cell, the investigator is interested on the number of plants cultivated subject to factors influencing its cultivation (see Kruijer et al. (2007) and references therein) and along with how the plants in the different neighbouring cells impact on the harvest in the (i, j)th position. Similarly in epidemiology, researchers are often concerned on the factors influencing the number of infected or death cases as a result of an outbreak of a virus in a region and how this is affecting the number of cases in the neighbouring regions. Such data has been treated in Cressie and Chan (1989), Wakefield (2007) while some more examples can be found in Lawson et al. (1999) and Lawson and Williams (2001). Moreover, in environmental field, the occurrences of road traffic accidents at different segments also illustrate spatial data analysis. In fact, in a hotspot analysis conducted in Barcelona, it was shown, via the local Moran statistics, that road accidents are concentrated in close neighbouring areas that have a complex road network systems with large roundabouts (Alvarez, 2020). Some other related research include the works by Valverde and Jovanis (2008) and Satria and Castro (2016). Last but not least, Mburu and Bakillah (2016) reported on the number of vehicle burglary incidents in small neighbouring regions of London which were highly spatially autocorrelated. Their study also revealed several influential factors such as unemployment and crimes in these areas of London.

Unsurprisingly, there is influx of models for the areal-type spatial data that include mainly the class of conditional autoregressive (CAR) models (Besag, 1974) and its extensions to Intrinsic CAR (ICAR) (Besag and Kooperberg, 1995), the Besag-Yorke-

Mollie (BYM) (Besag, Mollié and York, 1991) models and among several other extended CAR-based models; a review can be found in Obaromi (2019).

In fact, for the regular lattice data of discrete nature that are represented in terms of the site coordinates, the only works appearing in the literature so far use the spatial integer-valued auto-regressive model of order 1 (SINAR(1)) by Ghodsi, Shitan and Bakouch (2012) and Ghodsi (2015). The model was constructed by introducing dependence between the observation of interest with its unilateral spatial neighbouring observations via the binomial thinning operator defined by Steutel and Harn (1979). The structure is similar to the observation-driven integer-valued time series models defined in McKenzie (1986). Properties of the model including asymptotic properties of the CML estimators were thereon established and the spatial process was proven to be stationary and ergodic.

In the present paper we extend the model in certain directions. Firstly, we introduce different distributions for the innovations to enlarge the model and allow for larger variance, usually observed in spatial data due to clustering effects. Secondly, we allow for further spatial information to be used in the form of covariates that affect the model, leading to a non-stationary model. For this new model we discuss inference based on the CML. Moreover we discuss and apply some models related to the basic one that are parsimonious and easier to interpret, while they allow for easier extension to a broader family of models. Throughout the paper, some computational issues arising are also discussed.

The remaining of the paper proceeds as follows: The basic model and its extensions are described in Section 2. Simulations to further support the approach are provided in Section 3. A real data application related to the new models is provided in Section 4. Extensions of the current model and concluding remarks can be found in Section 5.

2 Generalised SINAR(1) model (GSINAR(1))

We consider spatial processes defined on a regular rectangular grid in two dimensions with sites labelled (i, j), with an associated random variable Y_{ij} defined at each site. Examples of such phenomena include data collected on a regular grid of size $n_1 \times n_2$ from satellites and from agricultural field trials. The unilateral model (see, e.g. Tjøstheim, 1978b, 1983; Basu and Reinsel, 1993) defines the neighbouring sites that provide information for the site (i, j), namely we denote as S_{ij} the set of indices (k, ℓ) of sites that are considered as neighbours of the site (i, j) and we define this as

$$S_{ij} = \{(k,\ell) \in Z^2 : k \leq i, \ell \leq j\} - \{(i,j)\}.$$

Tjøstheim (1983) described the model of order (p_1, p_2) for continuous data as:

$$Y_{ij} = \sum_{k=0}^{p_1} \sum_{\ell=0}^{p_2} \phi_{k\ell} Y_{i-k,j-\ell} + \epsilon_{ij},$$

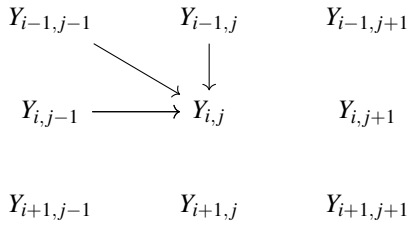

Figure 1: *A diagram representing the unilateral model. The arrows indicate which site influence the site under consideration. Only sites located at the left and upper from the site into consideration influence the site.*

where $\phi_{00} = 0$ and the errors ϵ_{ij} follow a $N(0,\sigma^2)$ distribution. The model of order $p_1 = p_2 = 1$ is described in Basu and Reinsel (1993). In the present paper we restrict our interest on this spatial structure, i.e. only of order 1. The assumed structure can be seen in Figure 1. One can see that for each point, only three neighbouring points coming from the upper left direction affect the point, implying a restricted spatial structure. In other words the spatial effect is propagated from left to the right and from top to the bottom only.

2.1 The stationary model

Ghodsi et al. (2012) extended the model for count data. Hence, they defined that the spatial observation located at site coordinate (i, j) follows an auto-regressive equation of the form:

$$Y_{ij} = \alpha_1 \circ Y_{i-1,j} + \alpha_2 \circ Y_{i,j-1} + \alpha_3 \circ Y_{i-1,j-1} + \epsilon_{ij}, \tag{1}$$

where $i = 1,\ldots,n_1$, $j = 1,\ldots,n_2$.

The dependence of Y_{ij} on its neighbours as defined by set S_{ij} is handled in equation (1) through the binomial thinning operator "$\circ$". The binomial thinning mechanism emanates from the work of Steutel and Harn (1979) (see also Scotto, Weiß and Gouveia, 2015) for a summary of such operators) and is expressed as:

$$\alpha \circ Y = \sum_{s=1}^{Y} B_s(\alpha), \tag{2}$$

where $\alpha \in [0,1]$, and $B_s(\alpha)$, $s = 1,\ldots,Y$ are identically and independently distributed Bernoulli r.v with $P(B_s(\alpha) = 1) = 1 - P(B_s(\alpha) = 0) = \alpha$. In parsimony, we impose the assumption of independent thinning operator (Du and Li, 1991; Bu, McCabe and Hadri, 2006). In (1), $\{\epsilon_{ij}\}_{i=1,\ldots,n_1,j=1,\ldots,n_2}$ represents the corresponding innovation sequence of independent non-negative integer-valued random variables with finite mean λ_ϵ and finite variance τ_ϵ^2 and has a distributional form as $P_{\epsilon_{ij}}(\cdot)$. Furthermore, at any position (i, j), ϵ_{ij} is assumed to be independent of all $Y_{i-k,j-\ell}$. In this simple form, the model in (1) is stationary if

$$\alpha_1 + \alpha_2 + \alpha_3 < 1. \tag{3}$$

Properties of this SINAR(1) model as well as estimation can be found in Ghodsi et al. (2012) and Ghodsi (2015). The process in equation (1) is proven to be ergodic in Ghodsi (2015) and Markovian in Pickard (1980).

2.2 The non-stationary model

Here, we extend the model to the non-stationary case by allowing site specific covariates to influence the mean of the innovation process. We denote the mean and variance of the innovation term as λ_{ij} and $\tau^2_{\epsilon_{ij}}$ respectively. We further consider λ_{ij} as a function of some position-variant and invariant covariates i.e. $\lambda_{ij} = f(x_{ij}^{\intercal}\boldsymbol{\beta})$ with $x_{ij} = [x_{ij1}, \ldots, x_{ijp}]^{\intercal}$ and regression coefficients $\boldsymbol{\beta} = [\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_p]^{\intercal}$. Note that a log function is a standard choice for such cases leading to

$$\log \lambda_{ij} = x_{ij}^{\intercal}\boldsymbol{\beta}.$$

We name this model as Generalized SINAR of order 1 (GSINAR(1)).

From the following binomial thinning properties,

$$\begin{aligned} E(\alpha \circ Y) &= \alpha E(Y) \\ V(\alpha \circ Y) &= \alpha(1-\alpha)E(Y) + \alpha^2 V(Y) \\ Cov(\alpha_1 \circ Y_1, \alpha_2 \circ Y_2) &= \alpha_1\alpha_2 Cov(Y_1, Y_2), \alpha_j \in (0,1), j = 1,2 \end{aligned}$$

and then we get the unconditional expectation of Y_{ij} to be

$$E(Y_{ij}) = \mu_{ij} = \alpha_1\mu_{i-1,j} + \alpha_2\mu_{i,j-1} + \alpha_3\mu_{i-1,j-1} + \lambda_{ij}. \tag{4}$$

For the variance we get

$$\begin{aligned} V(Y_{ij}) = \sigma^2_{ij} = {} & \alpha_1(1-\alpha_1)\mu_{i-1,j} + \alpha_1^2\sigma^2_{i-1,j} \\ & + \alpha_2(1-\alpha_2)\mu_{i-1,j} + \alpha_2^2\sigma^2_{i,j-1} \\ & + \alpha_3(1-\alpha_3)\mu_{i-1,j-1} + \alpha_3^2\sigma^2_{i-1,j-1} \\ & + 2\alpha_1\alpha_2 Cov(Y_{i-1,j}, Y_{i,j-1}) \\ & + 2\alpha_1\alpha_3 Cov(Y_{i-1,j}, Y_{i-1,j-1}) \\ & + 2\alpha_2\alpha_3 Cov(Y_{i,j-1}, Y_{i-1,j-1}) + \tau^2_{\epsilon_{ij}}. \end{aligned} \tag{5}$$

By letting $\gamma(k,\ell) = Cov(Y_{i-k,j-\ell}, Y_{ij})$, we obtain a difference equation of the form:

$$\gamma(k,\ell) = \alpha_1\gamma(k-1,\ell) + \alpha_2\gamma(k,\ell-1) + \alpha_3\gamma(k-1,\ell-1). \tag{6}$$

Closed form expressions for the above moments are difficult to obtain under non-stationary conditions. Whilst, unless assuming weak-stationarity, that is, ϵ_{ij} has constant mean λ_ϵ and variance τ_ϵ^2, we obtain simple expression for the mean and variance with the covariances obtained by solving the difference equation $\gamma(k,\ell)$ using the approach in Basu and Reinsel (1993). However, the derivation of the covariance structure in equation (6) is not required in the estimation of the model parameters when using the conditional maximum likelihood equation illustrated as follows. Conditional on the neighbourhood S_{ij}, the probability mass function for the GSINAR(1) model is given by:

$$\begin{aligned} P(Y_{ij}|S_{ij}) = & \sum_{s_1=0}^{R_1}\sum_{s_2=0}^{R_2}\sum_{s_3=0}^{R_3} p_{\alpha_1}(s_1|Y_{i-1,j})p_{\alpha_2}(s_2|Y_{i,j-1})p_{\alpha_3}(s_3|Y_{i-1,j-1}) \\ & P_{\epsilon_{ij}}(Y_{ij}-s_1-s_2-s_3), \end{aligned} \tag{7}$$

where $R_1 = min\{Y_{i-1,j}, Y_{ij}\}$, $R_2 = min\{Y_{i,j-1}, Y_{ij}-s_1\}$ and $R_3 = min\{Y_{i-1,j-1}, Y_{ij}-s_1-s_2\}$ and $p_\alpha(s|Y) = \binom{Y}{s}\alpha^s(1-\alpha)^{Y-s}$; $s = 0,1,\ldots,Y$, i.e. the probability mass function of a binomial distribution. In the present paper we have considered different choices for $P_{\epsilon_{ij}}(\cdot)$. The standard assumption of a Poisson distribution limits the variability we expect in the data (see Appendix). A natural improvement is to consider mixed Poisson alternatives. We consider the negative binomial, Poisson-Inverse Gaussian as well as the Poisson-Lindley in order to allow for quite different effects. Also, we consider the COM-Poisson distribution in order to allow for a general model which accounts for underdispersion if we need so. We postpone the details until section 3.

Therefore, the log conditional maximum likelihood (CML) equation is then given by:

$$\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=2}^{n_1}\sum_{j=2}^{n_2} \log P(Y_{ij}|S_{ij}), \tag{8}$$

where $\boldsymbol{\theta} = [\alpha_1, \alpha_2, \alpha_3, \boldsymbol{\beta}, \nu]$, ν refers to the dispersion parameter of the innovation distribution if it exists and $\boldsymbol{\beta}$ is the vector of regression coefficients for the mean of the innovation. It can be seen in (Ghodsi, 2015) that

$$\hat{\theta} - \theta \sim N(0, I^{-1}(\theta)),$$

where $I(\theta)$ is the Hessian matrix. The CML equation in (8) is then maximized.

Some computational details are the following. We have used the `optim` function in `R`. Note that the conditional distribution needs to derive the convolution of three binomials plus the distribution of the innovation term. This can be computationally intensive. We have reduced the computational burden by observing that the probabilities of the binomial distribution are just the coefficients of a polynomial of order N where N is the number of trials in the binomial. As such computing the convolution of two binomial is equivalent to multiply two polynomials for which there are very fast procedures, like those in the library `pracma` in `R`. This reduced the computational effort and improved with respect to the errors produces by huge finite summations. Overall, maximization of (8) was rather simple even for complicated innovation distributions.

2.3 Related models

The general model in (1) has three parameters to introduce the spatial correlation, namely α_1, α_2 and α_3 that described the vertical, horizontal and diagonal dependence respectively. One may eliminate some of the effects by setting the corresponding α parameter equal to 0. For example setting $\alpha_3 = 0$ we assume no-upper diagonal effect, while setting $\alpha_2 = 0$ we assume no horizontal effect. Such submodels can be very useful in order to examine and interpret the underlying situation for a dataset. For example, we may test and recognize which effect the vertical or horizontal is more important.

Another way to simplify the model is by assuming one common effect using the same parameter α for all directional relationships. Such a model takes the following form due to the properties of thinning operators:

$$Y_{ij} = \alpha \circ \sum_{(k,\ell)\in S_{ij}} Y_{k,\ell} + \epsilon_{ij}. \tag{9}$$

The model assumes that all neighbouring sites contribute the same to the structure. Such a model resembles simple INAR(1) time series models. It has the advantage of having less parameters to estimate and explain; all neighbours contribute the same. On the other hand, this may be restrictive since the spatial effects may differ due to direction and thus the model may fail to capture them correctly.

Model (9) allows for easy extensions to a general neighbouring structure. It is evident that by considering the set S_{ij} defining the neighbouring sites, this model can be generalized to a large extend including the non-regular lattice case which is more realistic in many applications. The model just assumes that all neighbours contribute to the observation at hand. Properties of such models as well as estimation is straightforward based on the results of the current section.

Finally, in the present paper we assume that the covariates enter in the model by the mean of the innovations. One may consider that spatial correlation parameters α_j may relate to some covariate information through a logit link function. For example, we can assume that $\text{logit}(\alpha_{1ij}) = x_{ij}^{\top}\boldsymbol{\delta}_1$, where x_{ij} is some covariate information for the site (i, j) and $\boldsymbol{\delta}_1$ some vector of regression coefficients. In this case we assume that each point in space has a different spatial effect α_1 depending on some characteristics x_{ij}. For example, we may assume that altitude can change the spatial effect, which makes sense if we measure for example something which can be altered due to wind conditions. We believe that such a model, while it has some potential is special cases, it can complicate the model interpretation, especially if we have regression effects in both the mean and the autocorrelation parameters.

3 Simulation study

In this section, we perform simulation experiments, using equation (1), with different innovation distributions namely the Poisson, Negative Binomial (NB), Poisson-Lindley (PL), Conway-Maxwell Poisson (COM-Poisson / CMP) and Poisson-Inverse Gaussian (PIG) with rate or mean parameters commonly indicated by a link predictor $\lambda_{ij} = \exp(x_{ij}^{\intercal}\boldsymbol{\beta})$ and dispersion index ν. As it is described in more details in the Appendix, using a distribution for the innovations that allows over/under dispersion, we also extend such properties to the observed spatial distribution and hence more realistic and flexible fitting can be achieved. In the present paper we consider the following distributions for the innovations.

- For Poisson innovations we assume that

$$P(\epsilon_{ij}) = \frac{e^{-\lambda_{ij}}\lambda_{ij}^{\epsilon_{ij}}}{\epsilon_{ij}!}; \epsilon_{ij} = 0,1,\ldots; \lambda_{ij} \geq 0.$$

- For NB innovations we use the following parametrization:

$$P(\epsilon_{ij}) = \frac{\Gamma(\nu+\epsilon_{ij})}{\Gamma(\nu)\epsilon_{ij}!}\left(\frac{\lambda_{ij}}{\lambda_{ij}+\nu}\right)^{\nu}\left(\frac{\nu}{\lambda_{ij}+\nu}\right)^{\epsilon_{ij}}; \epsilon_{ij} = 0,1,\ldots; \lambda_{ij} \geq 0, \nu \geq 0.$$

  For this parametrization the mean is λ_{ij} and the variance $\lambda_{ij} + \lambda_{ij}^2/\nu$.

- For PL innovations we use

$$P(\epsilon_{ij}) = \frac{\lambda^2(\epsilon_{ij}+\lambda_{ij}+2)}{(\lambda_{ij}+1)^{\epsilon_{ij}+3}}; \epsilon_{ij} = 0,1,\ldots; \lambda_{ij} > 0.$$

  The mean is $(\lambda_{ij}+2)/(\lambda_{ij}(\lambda_{ij}+1))$ while the variance is

$$\frac{\lambda_{ij}^3+4\lambda_{ij}^2+6\lambda_{ij}+2}{\lambda_{ij}^2(\lambda_{ij}+1)^2}$$

  PL can have different shapes than the other Poisson mixtures like the NB and PIG models.

- For COM-Poisson innovations we use

$$P(\epsilon_{ij}) = \frac{\lambda_{ij}^{\epsilon_{ij}}}{(\epsilon_{ij})!^{\nu}}\frac{1}{Z(\lambda_{ij},\nu)}; \epsilon_{ij} = 0,1,\ldots; j = 0,1,\ldots; \lambda_{ij} \geq 0; \nu \geq 0,$$

  where $Z(\lambda_{ij},\nu) = \sum_{j=0}^{\infty}\frac{\lambda_{ij}^j}{j!^{\nu}}$. Note that λ_{ij} is not the mean of the distribution; the mean is hard to be written in closed form, but it is approximated by $\lambda_{ij}^{1/\nu}$.

- For PIG innovations we use the parameterization from the package `actuar` in `R`. Namely the pmf is defined with parameters λ and dispersion ν as

$$P(\epsilon_{ij}) = \left(\frac{2\nu}{\pi}\right)^{1/2} \exp\left(\frac{\nu}{\lambda_{ij}}\right) \left(\frac{a_{ij}}{\nu}\right)^{-(x-\frac{1}{2})} K_{x-\frac{1}{2}}(a_{ij}),\ \ x = 0,1,\ldots,\ \ \lambda_{ij},\nu > 0,$$

where $a_{ij}^2 = 2\nu\left(1+\frac{\nu}{2\lambda_{ij}^2}\right)$ and $K_x(a)$ is the modified Bessel function of the third kind. The mean is λ_{ij} and the variance is $\lambda_{ij} + \lambda_{ij}^3/\nu$.

The choice of the distributions for the innovation term attempts to cover a range of possible models. So, we have used the Poisson distribution as a starting point, two of the most famous mixed Poisson ones (negative binomial and Poisson-Inverse Gaussian), the COM-Poisson to allow for under-dispersion as well and finally the Poisson-Lindley since this is a tractable mixed Poisson distribution with very different shapes.

In order to simulate the grid we followed the following approach: We added an additional row and column with all values equal to 0, i.e. we set $Y_{0j} = Y_{i0} = 0$ for all i and j. Then we simulated the grid Y_{ij}, $i = 1,\ldots,n_1+10$ and $j = 1,\ldots,n_2+10$ based on the model in (1) using the chosen innovation distribution. Then we rejected the rows and columns from 0 up to 10 so as to keep the grid $n_1 \times n_2$.

3.1 Numerical Results: No covariates

For scenario 1, the simulation study assumes the following combinations of $(\alpha_1, \alpha_2, \alpha_3, \lambda, \nu)$ and grids:

1. C1: (0.35, 0.15, 0.2, 5, 0.5) and grid 25 x 25
2. C2: (0.25, 0.25, 0.3, 3, 0.8) and grid 40 x 40
3. C3: (0.6, 0.2, 0.15, 7, 2) and grid 50 x 50

Note that for Poisson-Lindley and COM-Poisson cases parameter λ is not the mean while ν has a different interpretation. So, in the simulations this is the value used to simulate the data. For Poisson, negative binomial and Poisson-Inverse Gaussian, λ is the mean and ν is the dispersion parameters, equal to 1 for the Poisson. Obviously for $\nu \to \infty$ we get the Poisson distribution in such cases.

For each scenario 1000 replications were obtained. The simulated mean estimates, their biases, root mean square errors (RMSEs) and standard deviations (SDs) are reported. The results are displayed in Tables 1, 2 and 3.

Table 1: *Mean, Bias, RMSE and SD of estimates under different innovations under C1. Note that for Poisson-Lindley and COM-Poisson case parameter λ is not the mean. In the simulations this is the value used to simulate the data.*

Innovation	Estimates	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\alpha_3}$	$\hat{\lambda}$	$\hat{\nu}$
Poisson	Mean	0.3503	0.1485	0.2018	4.9558	
	Bias	0.0003	−0.0015	0.0018	−0.0442	
	RMSE	0.0360	0.0425	0.0438	0.7580	
	SD	0.0350	0.0405	0.0338	0.7571	
NB	Mean	0.3555	0.1502	0.2178	5.5005	0.5660
	Bias	0.0055	0.0002	0.0178	0.5005	0.0660
	RMSE	0.0521	0.0601	0.0489	0.7280	0.4180
	SD	0.0355	0.0436	0.0426	0.6597	0.5296
PL	Mean	0.3477	0.1466	0.1969	5.0892	
	Bias	−0.0023	−0.0034	−0.0031	0.0892	
	RMSE	0.0367	0.0538	0.0715	0.7141	
	SD	0.0369	0.0541	0.0719	0.7233	
CMP	Mean	0.3557	0.1478	0.2035	4.2440	0.5770
	Bias	0.0057	−0.0022	0.0035	−0.7560	0.0770
	RMSE	0.0385	0.0502	0.0485	0.8524	0.5912
	SD	0.0381	0.0402	0.0414	0.6915	0.5996
PIG	Mean	0.3414	0.1416	0.2101	4.8164	0.4990
	Bias	−0.0086	−0.0084	0.0101	0.1836	−0.0010
	RMSE	0.0345	0.0350	0.0490	0.7839	0.5813
	SD	0.0399	0.0111	0.0261	0.6781	0.5793

Table 2: *Mean, Bias, RMSE and SD of estimates under different innovations under C2. Note that for Poisson-Lindley and COM-Poisson case parameter λ is not the mean. In the simulations this is the value used to simulate the data.*

Innovation	Estimates	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\alpha_3}$	$\hat{\lambda}$	$\hat{\nu}$
Poisson	Mean	0.2500	0.2489	0.2989	3.0290	
	Bias	0.0000	−0.0011	−0.0011	0.0290	
	RMSE	0.0010	0.0009	0.0011	0.0172	
	SD	0.0008	0.0007	0.0010	0.0152	
NB	Mean	0.2497	0.2505	0.3012	3.0389	0.7990
	Bias	−0.0003	0.0005	0.0012	0.0389	−0.0010
	RMSE	0.0218	0.0238	0.0294	0.6433	0.2543
	SD	0.0118	0.0218	0.0154	0.6354	0.2891
PL	Mean	0.2491	0.2479	0.2999	2.9990	
	Bias	−0.0009	−0.0021	−0.0001	−0.0010	
	RMSE	0.0216	0.0216	0.0245	0.2654	
	SD	0.0206	0.0211	0.0204	0.2655	
CMP	Mean	0.2487	0.2481	0.3009	3.0266	0.8009
	Bias	−0.0023	−0.0019	0.0009	0.0266	0.0009
	RMSE	0.0223	0.0227	0.0239	0.4625	0.4728
	SD	0.0222	0.0225	0.0209	0.4619	0.4731
PIG	Mean	0.2480	0.2488	0.2979	2.9901	0.8111
	Bias	−0.0020	−0.0012	−0.0021	−0.0099	0.0111
	RMSE	0.0183	0.0176	0.0210	0.3014	0.4467
	SD	0.0182	0.0106	0.0209	0.3004	0.4807

Table 3: *Mean, Bias, RMSE and SD of estimates under different innovations under C3. Note that for Poisson-Lindley and COM-Poisson case parameter λ is not the mean. In the simulations this is the value used to simulate the data.*

Innovation	Estimates	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\alpha_3}$	$\hat{\lambda}$	$\hat{\nu}$
Poisson	Mean	0.5997	0.2070	0.1495	6.9875	
	Bias	−0.0003	0.0070	−0.0005	−0.0125	
	RMSE	0.0009	0.0003	0.0010	0.0101	
	SD	0.0005	0.0003	0.0009	0.0100	
NB	Mean	0.5996	0.2004	0.1511	6.9973	2.0010
	Bias	−0.0004	0.0004	0.0011	−0.0027	0.0010
	RMSE	0.0211	0.0120	0.0151	0.0910	0.2170
	SD	0.0218	0.0140	0.0170	0.0987	0.2281
PL	Mean	0.6067	0.2016	0.1503	6.9898	
	Bias	0.0067	0.0016	0.0003	−0.0102	
	RMSE	0.0207	0.0140	0.0133	0.0119	
	SD	0.0195	0.0134	0.0123	0.0102	
CMP	Mean	0.5982	0.1918	0.1529	7.0131	2.0210
	Bias	−0.0018	−0.0082	0.0029	0.0131	0.0210
	RMSE	0.0210	0.0130	0.0214	0.3798	0.3720
	SD	0.0214	0.0134	0.0215	0.3898	0.3731
PIG	Mean	0.5991	0.1992	0.1499	6.9997	1.9830
	Bias	−0.0009	−0.0008	−0.0001	−0.0003	−0.0170
	RMSE	0.0151	0.0116	0.0150	0.0998	0.5430
	SD	0.0152	0.0115	0.0154	0.0950	0.4450

The simulation results illustrate that the estimates of the different parameters are consistent. This remark is noticed for all the SINAR with the different innovation distributions and under the different combinations of C1, C2, C3. The simulations also ensured that the estimates of the $\hat{\alpha}$'s satisfy the stability condition for stationarity given in (3).

Note that in all replications almost no problems to maximize the log-likelihood were detected. Some problems occurred in the COM-Poisson innovations. Problems are related to the built in functions `dcomp` and `dcompoisson` as they could not compute efficiently the normalizing constant $Z(\lambda,\nu)$ in the COM-Poisson implementations in few simulations.

3.2 Numerical Results: With covariates

For the case with covariates we have added a covariate, say, X for the different scenarios. So we assume for the innovations that

$$\log \lambda_{ij} = \beta_0 + \beta_1 X_{ij},$$

where the covariate X_{ij} was generated from a standard normal distribution. Again we have checked different grids, namely 30×30, 50×50 and 80×80 to see how the size

of the grid scales up the variance and the biases (if any). We have used two scenarios:

- S1: $\alpha_1 = 0.15, \alpha_2 = 0.1, \alpha_3 = 0.2, \beta_0 = 0.6, \beta_1 = 0.5$,
- S2: $\alpha_1 = 0.05, \alpha_2 = 0.1, \alpha_3 = 0.05, \beta_0 = 0.1, \beta_1 = -0.5$.

One can see that the second scenario S2 has smaller spatial correlation parameters closer to the lower boundary and hence we would like to see the behaviour. Now we need to estimate all 5 parameters.

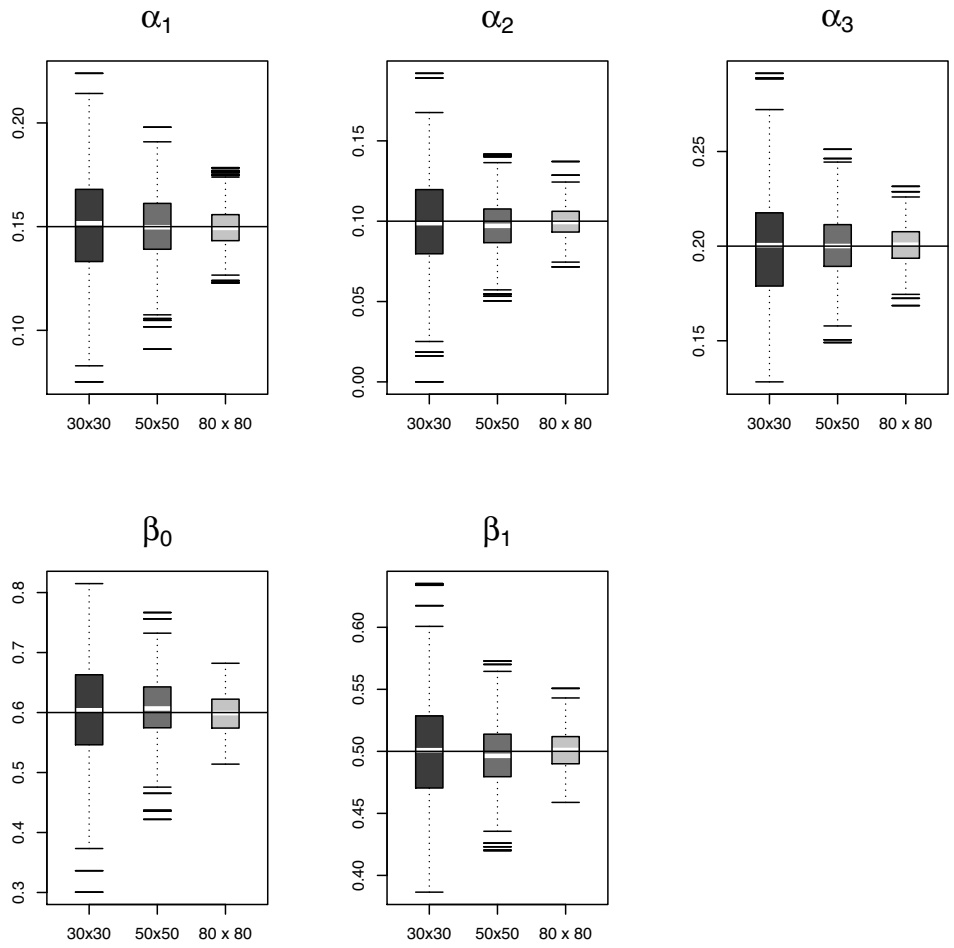

Figure 2: *Boxplots for all five parameters under scenario S1 for the three different grids. The horizontal line represents the true value.*

Figures 2 and 3 show the boxplots from 1000 replications under the two scenarios for the different grids. We present results from the Poisson innovations case only and similar

findings were obtained from the other models as well. The horizontal line represents the true value. One can see that even in the smaller grid the CML estimates correctly the true value. The variability as indicated by the boxplots reduces with the grid size as expected. Also from the boxplots one can see that the shape is symmetrical and confront with the asymptotic normality of the estimates.

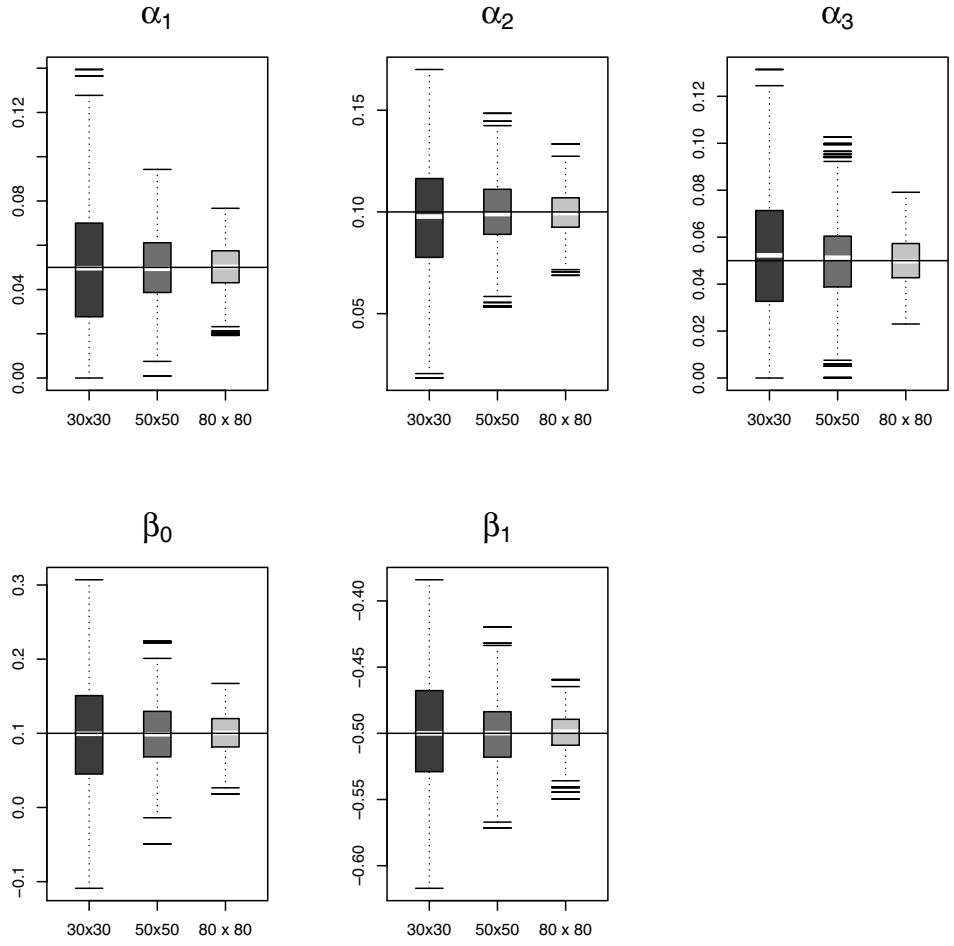

Figure 3: *Boxplots for all five parameters under scenario S2 for the three different grids. The horizontal line represents the true value.*

As far as the computational issues are concerned, no problems were found when fitting the models. As initial values we have used random value around the true ones, i.e. we simulated initial values by adding a uniform random variable in the interval $(-0.05, 0.05)$ to the true underlying values. For all runs we got convergence from `optim` function. To obey the restrictions of the parameter space we used transformations on the parameters.

4 Applications - Beilschmiedia data

4.1 The data

In studies of biodiversity of tropical rainforests, it is of interest to study whether the spatial patterns of the many different tree species can be related to spatial variations in environmental variables concerning topography and soil properties. Beilschmiedia dataset (Bei) in the `spatstat` package in `R` (Baddeley, Rubak and Turner, 2015) captures the locations of 3605 trees in a tropical rain forest . The data cover a 1000 m $\times$ 500 m rectangular sampling region in the tropical rainforest of Barro Colorado Island. This data set is a part of a much larger data set containing positions of hundreds of thousands of trees belonging to hundreds of species. More details about the data can be found in Møller and Waagepetersen (2007).

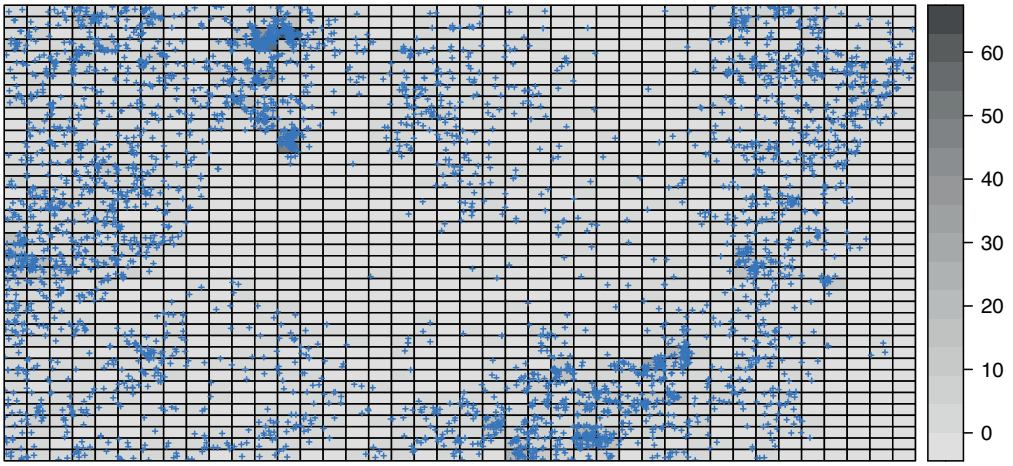

Figure 4: *Spatial plot of Bei data.*

A regular lattice of size 40 $\times$ 40 is created from the original dataset considering the number of trees inside each cell. The mean and variance are 2.25 and 12.4 respectively. The index of dispersion is 5.51 implying overdispersion that we cannot capture by Poisson innovations. Fitting models that allow for overdispersion is important. Figure 4 represents a spatial plot of the Bei data. In particular one can see the position of the trees in the lattice. The grey background depicts the observed counts and darker grey areas are those with more trees. In addition, Table 4 shows some values of the sample spatial autocorrelation of order k and ℓ for the Bei data. Here k refers to the horizontal direction and ℓ to the vertical direction. One can see that the horizontal autocorrelations are larger, supporting the use of a model like the one derived in section 2. The observed counts in the 40×40 grid can be also seen in Figure 5.

Table 4: *Some values of the sample spatial autocorrelation for the Bei data.*

			k		
ℓ	0	1	2	3	4
0	1.000	0.497	0.238	0.283	0.269
1	0.379	0.265	0.223	0.228	0.170
2	0.172	0.125	0.104	0.133	0.132
3	0.094	0.090	0.083	0.054	0.049
4	0.083	0.071	0.046	0.039	0.023

4.2 Results

To start with, we have fitted a series of models with different innovation distributions to capture the different aspects of the data. The fitted model to the Bei data using the CML estimation approach can be seen in Table 5.

Table 5: *Comparison between different innovation distributions using Akaike information criterion (AIC): Application to BEI count data.*

Innovations		$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\alpha_3}$	$\hat{\lambda}$	$\hat{\nu}$	AIC
Poisson	Estimates	0.198	0.306	0.108	0.818		
	s.e	(0.019)	(0.013)	(0.011)	(0.034)		7179
NB	Estimates	0.151	0.225	0.053	1.209	0.205	
	s.e	(0.056)	(0.017)	(0.015)	(0.082)	(0.018)	5342
PL	Estimates	0.119	0.223	0.008	1.070		
	s.e	(0.016)	(0.017)	(0.015)	(0.039)		5882
CMP	Estimates	0.113	0.212	0.002	0.594	0.001	
	s.e	(0.017)	(0.018)	(0.015)	(0.028)	(0.042)	5732
PIG	Estimates	0.151	0.223	0.060	1.214	0.161	
	s.e	(0.015)	(0.016)	(0.015)	(0.097)	(0.022)	5343

It is observed from Table 5 that the SINAR(1) model using NB innovation distribution yields the lowest AIC value and hence outperforms the other models with different innovations. However, an interesting observation is that the PIG model has an AIC value which is really close to the selected model. In fact this implies that we need an overdispersed innovation distribution to capture the observed overdispersion. Note also that for all models the dependence parameters α_j are significant, supporting the usage of spatial models. We can also observe that the α_j are all positive values showing that geographically nearby values of the variable of interest are more similar than those of remote locations. Parameter α_2 that measures the horizontal dependence is larger. Perhaps this may relate to parameters associated with the lattice like the orientation with respect prevailing winds that expand the vegetation to some particular direction. For the COM-Poisson distribution, the model tends to a geometric distribution since parameter ν is almost zero. This may explain why the fit is not that good.

Table 6: *Different models with negative binomial innovation distribution fitted to the Bei data. Models assumes different (or not ta all) spatial dependence. The full model with all three kind of unilateral effects is the chosen one.*

Model	Param	Log-lik	AIC
No restriction	5	−2666.22	5337.442
M1: $\alpha_1 = \alpha_2 = \alpha_3$	3	−2686.20	5375.406
M2: $\alpha_3 = 0$	4	−2674.75	5353.490
M3: $\alpha_2 = 0$	4	−2765.51	5535.024
M4: $\alpha_1 = 0$	4	−2717.78	5439.554
M5: $\alpha_2 = \alpha_3 = 0$: only verctical	3	−2802.29	5607.586
M6: $\alpha_1 = \alpha_3 = 0$: only horizontal	3	−2746.06	5495.128
M7: $\alpha_1 = \alpha_2 = 0$: only diagonal	3	−2840.98	5684.954
no spatial effect $\alpha_1 = \alpha_2 = \alpha_3 = 0$	2	−2938.80	5879.606

Table 6 presents for the chosen negative binomial case some more spatial scenarios as mentioned in section 2.3. For example, model M1 assumes that all three α's are the same, while models M2 to M4 that one of the spatial correlations is not present i.e. we set $\alpha_j = 0$ for $j = 1,2,3$ respectively. Actually we remove each time the vertical (M4), horizontal (M3) and the diagonal effects (M2). Finally, models M5 to M7 suggest that only one spatial effect suffices. The full model with all three kind of unilateral effects is the chosen one as judged by AIC, revealing the underlying structure of the data. One can see that the horizontal effect is larger as judged by the change in the LRT when we remove each effect.

In addition we use some covariate information available. The Bei data set is accompanied by covariate data giving the elevation (altitude) and slope of elevation in the study region. An important question arises is whether the intensity of Bei trees may be viewed as a spatially varying function of the covariates. We have fitted different models to examine the improvement offered by the covariates. Both covariates were found significant. The selected model can be seen in Table 7.

Table 7: *Comparison between Poisson innovation and mixed Poisson innovations using Akaike information criterion (AIC): Application to Bei count data with both covariates.*

Innovations		$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\alpha_3}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\nu}$	AIC
Poisson	Estimates	0.187	0.286	0.102	-6.265	0.037	8.128		
	s.e	(0.012)	(0.013)	(0.011)	(0.671)	(0.004)	(0.478)		6940
NB	Estimates	0.146	0.219	0.055	-6.897	0.044	8.515	0.242	
	s.e	(0.016)	(0.017)	(0.016)	(1.507)	(0.010)	(1.327)	(0.089)	5293
PL	Estimates	0.115	0.208	0.000	4.904	-0.029	-6.839		
	s.e	(0.016)	(0.017)	(0.014)	(0.645)	(0.004)	(0.546)		5716
CMP	Estimates	0.111	0.202	0.006	-2.162	0.010	2.089	0.001	
	s.e	(0.016)	(0.018)	(0.015)	(0.116)	(0.001)	(0.236)	(0.031)	5626
PIG	Estimates	0.146	0.221	0.056	-15.053	0.097	16.539	0.210	
	s.e	(0.016)	(0.017)	(0.015)	(2.395)	(0.016)	(3.451)	(0.0213)	5294

It is observed from Table 7 that the SINAR(1) model using NB innovation distribution with covariates still yields the lowest AIC value and hence outperforms the other models with different innovations by producing much better fit to the data. PIG is very competitive to the NB.

We focus on the selected model with negative binomial innovations. From the results of Table 7 one can see that both covariates are statistically significant, with positive sign, hence increasing the altitude and the slope of elevation we obtain an increased number of trees, having adjusted for the effect of neighbouring areas. From the α's we see that the larger effect comes from α_2 that measures the horizontal dependence. All spatial effects are statistically significant at 5%. The model implies a clear spatial dependence.

4.3 Goodness of Fit

In order to jude whether the fitted model is satisfactory we have worked a few ideas. To start with, we derived the one step ahead predictions based on the models. Namely, we derived for each data point

$$E(Y_{ij}|S_{ij}) = \hat{\alpha}_1 Y_{i-1,j} + \hat{\alpha}_2 Y_{i,j-1} + \hat{\alpha}_3 Y_{i-1,j-1} + \hat{\lambda}_{ij}$$

and

$$\log \hat{\lambda}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 X_{1ij} + \hat{\beta}_2 X_{2ij}$$

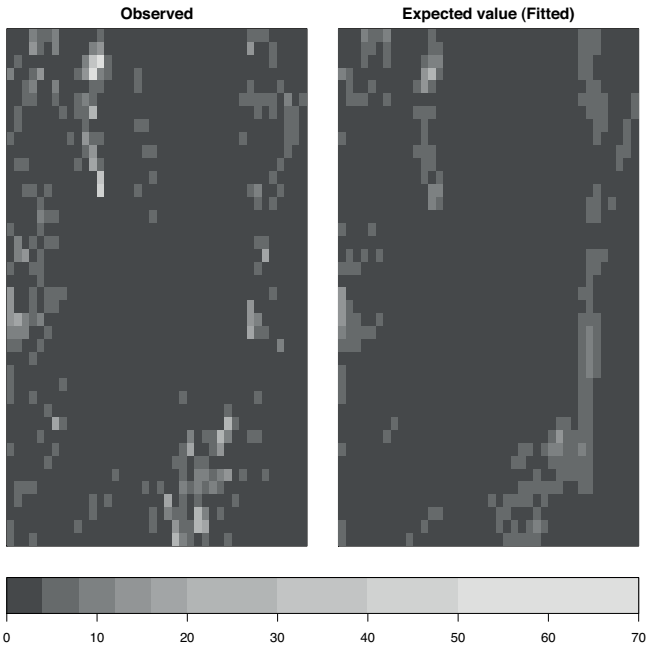

Figure 5: *The observed counts in the* 40×40 *grid and the fitted based on the model. Fitted values are expected values based on the conditional expectation.*

using the estimated parameters from the selected negative binomial model. The values can be seen in Figure 5 together with the observed counts. We emphasize that the predictions are the expected means that is why they cannot capture the extreme values. From the plots one can see that the model captures in a great extend the pattern.

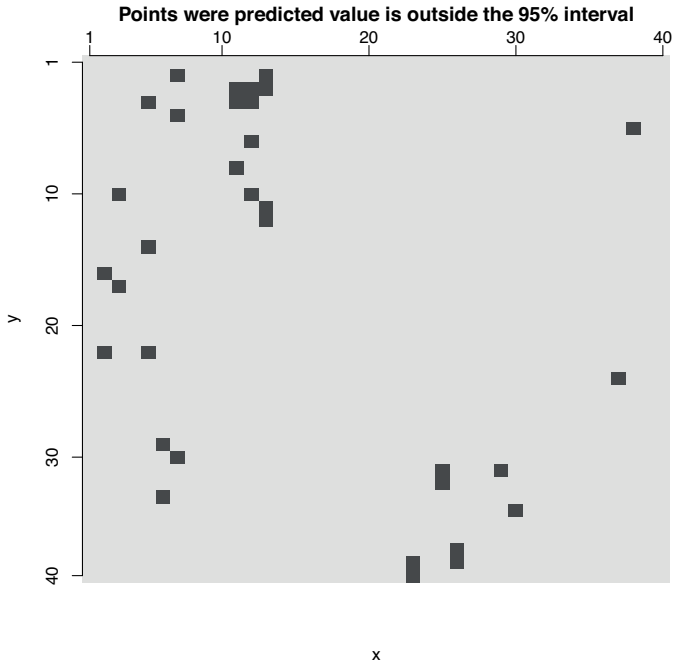

Figure 6: *Points in the grid that the true value was outside the 95% prediction interval created.*

To further exploit the quality of the predictions we have created for each data point (i, j) a 95% confidence interval for the prediction. To do so, we simulated 1000 values from the predictive probability mass function as provided in (7) and then based on them we created the intervals. Out of the 1521 values we predict (we did not predict the first row and column) only 33 (2.1%) values were outside of the interval, implying that the model was quite satisfactory. The values that lay outside the interval are depicted in Figure 6. One can see again that we have failed to predict some of the extreme values as one can see compared to the Figure 5.

Another important aspect of the model fitting lies on the ability of the model to capture the spatial dependence structure. To check this aspect we simulated grids of the same size 40×40 from the selected negative binomial using the estimated parameters. For each simulation we have estimated the spatial covariance at lags k and ℓ by

$$\hat{\gamma}(k,\ell) = \frac{1}{n} \sum_{i=k+1}^{40} \sum_{j=\ell+1}^{40} (Y_{ij} - \bar{Y})(Y_{i+k,j+\ell} - \bar{Y})$$

and then we derived the spatial correlation at lag k and ℓ as

$$\hat{\rho}(k,\ell) = \frac{\hat{\gamma}(k,\ell)}{s_Y^2},$$

where $\bar{Y}$ and s_Y^2 are the mean and the variance estimated from the data.

We have used lags $k = 0,1,2$ and $\ell = 0,1,2$ and then we compared them with those values observed from the data in order to see whether the observed dependence structure could have been created by the model at hand. We show in Figure 7 the 95% confidence intervals created by 1000 simulations and the dot indicates the observed value. We see a good agreement. The two first values need perhaps improvement with a richer model but overall the model captures the underlying structure in a reasonable way.

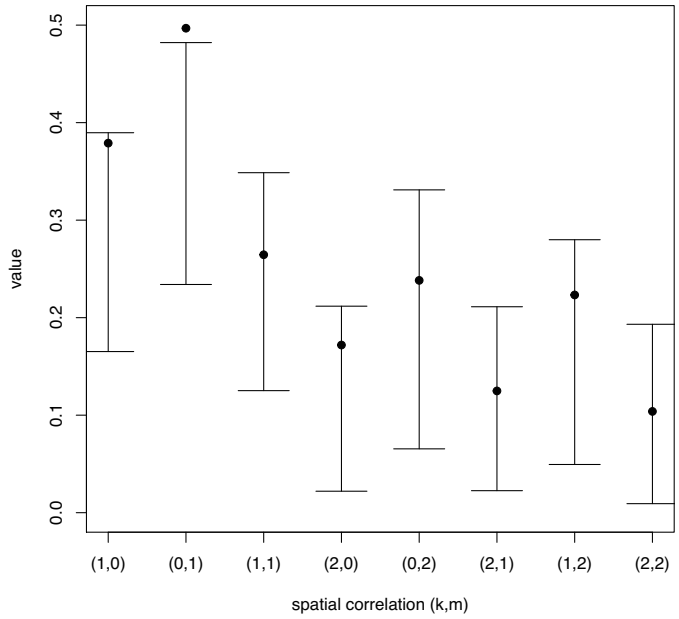

Figure 7: *Points in the grid that the true value was outside the 95% prediction interval created.*

As a summary, we believe that the current model fits satisfactorily the observed spatial structure.

5 Conclusion

This paper revisits and extends the simple stationary SINAR(1) model introduced by Ghodsi et al. (2012) and Ghodsi (2015). The SINAR(1) model is the first research in the modelling of the two dimensional unilateral spatial discrete data based on the thinning mechanism that allows to model explicitly the discrete nature of the data.

In the present paper we proposed some novel extensions of the existing SINAR(1) model. These novelties in fact overcome two important limitations of the simple SINAR(1). Firstly, in our model specification, we propose to model the data using overdispersed innovation distributions, while simultaneously allow covariate information to be used leading to a non-stationary model. While not treated in this version, one may also use offset in the regression part, like population size values, if needed. We also discuss parsimonious representations of the model at hand. The model parameters are estimated using the CML approach.

We acknowledge some restrictions of the current model which we consider to improve. The proposed model is based on the unilateral regular lattice case. One can extend the model to capture several other cases. For the regular lattice case, define the set of indices (k,ℓ) of the neighbouring observations for the (i,j) observation as S_{ij}. Then, in the general setting, the model can be written as for a general neighbourhood:

$$Y_{ij} = \sum_{(k,\ell)\in S_{ij}} \alpha_{k\ell} \circ Y_{k,\ell} + \epsilon_{ij},$$

where are usual, the ϵ_{ij} are the innovations. Defining appropriately the sets S_{ij} one can derive other models at the expense of parsimony.

Finally consider the typical case in spatial data where observations are indexed simply as Y_j to indicate the value at site j from a map with $j = 1,\ldots,n$ sites, as for example the different regions of a country. Define as S_j the indices of its neighbours. In such case the model of order 1 can have the form:

$$Y_j = \alpha \circ \sum_{k\in S_j} Y_k + \epsilon_j$$

or equivalently if we define the $n \times n$ adjacency matrix W with elements w_{ij} with values equal to 1 if the sites i and j are neighbours and 0 otherwise, then we can write the models as:

$$Y_j = \alpha \circ \sum_{k\neq j} w_{kj} Y_k + \epsilon_j$$

to mimic typical order 1 models for spatial continuous data. Such generalization will be reported elsewhere.

Also note that in this paper we used only spatial model of first order. One may consider SINAR(p) models with higher order effects. Such extension needs special care. It is already known that simple INAR(p) models can have different interpretations /representations, (see the different approaches in Alzaid and Al-Osh (1990) and Jin-Guan and Yuan (1991)). Extending to a SINAR(p) model can have a large number of parameters making inference quite complex. Perhaps more parsimonious models like the one in Section 2.3 are easier to extend to higher order.

A Appendix section

Based on Ghodsi et al. (2012), for the case of a stationary model, we have for the marginal stationary mean μ_Y and the stationary variance σ_Y^2 that

$$\mu_Y = \frac{\mu_\epsilon}{1-\alpha_1-\alpha_2-\alpha_3}$$

and

$$\sigma_Y^2 = \frac{\mu_Y \sum_{i=1}^{3} \alpha_i(1-\alpha_i) + \tau_\epsilon^2}{1-(\alpha_1+\alpha_2\alpha_3)\lambda-(\alpha_2+\alpha_1\alpha_3)\eta-\alpha_3^2}$$

where $\eta = \frac{\alpha_2+\alpha_3\lambda}{1-\alpha_1\lambda}$ and

$$\lambda = \frac{(1+\alpha_1^2-\alpha_2^2-\alpha_3^2) - \sqrt{(1+\alpha_1^2-\alpha_2^2-\alpha_3^2)^2 - 4(\alpha_1+\alpha_2\alpha_3)^2}}{2(\alpha_1+\alpha_2\alpha_3)}$$

In the formulas, μ_ϵ and τ_ϵ^2 are the mean and the variance of the innovations respectively. Note a misprint in Ghodsi et al. (2012) for the variance. Define the index of dispersion $ID_Y = \sigma_Y^2/\mu_Y$. Dividing the variance with the mean, we get for the index of dispersion of the spatial data that

$$ID_Y = \frac{\sigma_Y^2}{\mu_Y} = \frac{\sum_{i=1}^{3} \alpha_i(1-\alpha_i) + ID_\epsilon(1-\alpha_1-\alpha_2-\alpha_3)}{1-(\alpha_1+\alpha_2\alpha_3)\lambda-(\alpha_2+\alpha_1\alpha_3)\eta-\alpha_3^2}$$

which relates directly the index of dispersion of the innovation distribution ID_ϵ to that of the marginal, i.e. ID_Y. Since the denominator is positive and all the quantities in the nominator also are positive, an increase of ID_ϵ will lead to increase of the ID_Y. Thus assuming an overdispersed distribution for the innovations we can have much larger overdispersion in the observed spatial data.

One can see that even for the Poisson innovations the index of dispersion is larger than 1, however for reasonable values for counts this overdispersion is limited. The introduction of overdispersed innovations increase a lot the overdispersion as one can see in Figure 8. In Figure 8 the two axes depict the marginal mean and variance for a stationary model given above. The different lines correspond to different levels of overdispersion for the innovation distribution. We have used $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$. The diagonal line refers to the case of equidispersion. Therefore, above that line we get overdispersion and below underdispersion. The red line (ID=1) corresponds to Poisson innovations. One can see that in this case we get small overdispersion for the spatial case. Increasing the overdispersion on the innovation, as for example considering a mixture of Poisson we get larger overdispersion. Note that an underdispersed innovation distribution, like the cases of COM-Poisson distribution, can lead to underdispersion.

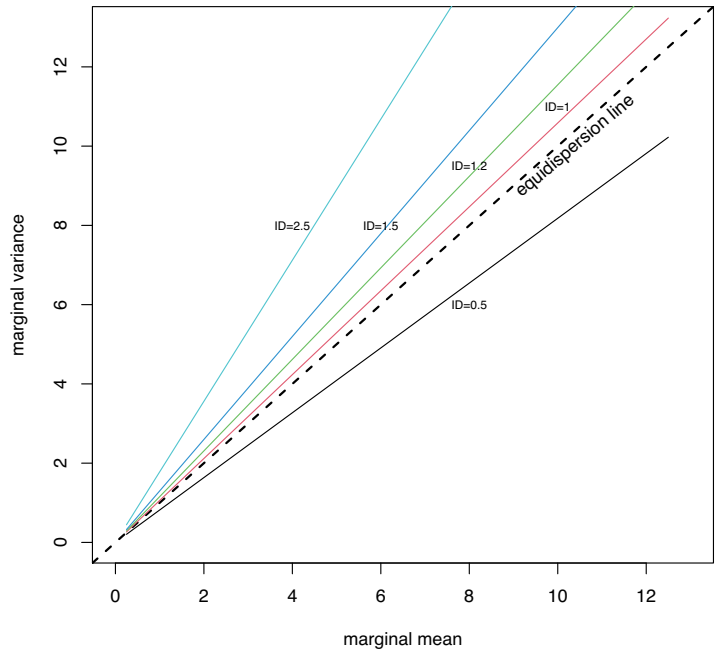

Figure 8: *The marginal mean and variance for a stationary model. The different lines correspond to different levels of overdispersion for the innovation distribution. for the plot* $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$. *The diagonal line refers to the case of equidispersion. ID implies the index of dispersion of the innovation distribution. One can see that for Poisson (red line) we get small overdispersion. Increasing the overdispersion of the innovations lead to increased overdispersion for the spatial distribution.*

References

Alvarez, M. (2020). Predicting traffic accident hotspots with spatial data science. https://carto.com/blog/predicting-traffic-accident-hotspots-with-spatial-data-science/

Alzaid, A. and Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, 314-324.

Baddeley, A., Rubak, E. and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with* R. London: Chapman; Hall/CRC press.

Basu, S. and Reinsel, G. C. (1993). Properties of the spatial unilateral first-order ARMA model. *Advances in Applied Probability*, 631-648.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B*, 36, 192-236.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82, 733-746.

Besag, J., Mollié, A. and York, J. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1-20.

Bu, R., McCabe, B. and Hadri, K. (2006). Maximum likelihood estimation of higher-order integer-valud autoregressive processes. *Journal of Time Series Analysis*, 29, 973-994.

Chun, Y. (2014). An application to vehicle burglary. *Geographical Analysis*, 46, 165-184.

Cressie, N. (1993). *Statisics for Spatial Data* (2nd ed.). New York: Wiley.

Cressie, N. and Chan, N. (1989). Spatial modeling of regional variables. *Epidemiology and Infection*, 84, 393-401.

Du, J. and Li, Y. (1991). The integer-valued autoregressive INAR(p) model. *Journal of Time Series Analysis*, 12, 129-142.

Ghodsi, A. (2015). Conditional maximum likelihood estimation of the first-order spatial integer-valued autoregressive SINAR(1,1) model. *Journal of The Iranian Statistical Society*, 14, 15-36.

Ghodsi, A., Shitan, M. and Bakouch, H. (2012). A first-order spatial integer-valued autoregressive SINAR(1,1) model. *Communications in Statistics- Theory and Methods*, 41, 2773-2787.

Jin-Guan, D. and Yuan, L. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, 12, 129-142.

Kruijer, W., Stein, A., Schaasfsma, W. and Heijting, S. (2007). Analyzing spatial count data, with an application to weed counts. *Environmental and Ecological Statistics*, 14, 399-410.

Lawson, A., Biggeri, A., Bohning, D., Lesaffre, E., Viel, J. and Bertollini, R. (1999). *Disease Mapping and Risk Assessment for Public Health*. Chichester, UK: Wiley.

Lawson, A. and Williams, F. (2001). *An Introductory Guide to Disease Mapping*. New York: Wiley Medical Sciences.

Mburu, L. and Bakillah, M. (2016). Modeling spatial interactions between areas to assess the burglary risk. *International Journal of Geo-Information*, 5, 47.

McKenzie, E. (1986). Autoregressive moving-average processes with negative binomial and geometric marginal distrbutions. *Advanced in Applied Probability*, 18, 679-705.

Møller, J. and Waagepetersen, R. (2007). Modern spatial point process modelling and inference (with discussion). *Scandinavian Journal of Statistics*, 34, 643-711.

Obaromi, D. (2019). Spatial modelling of some conditional autoregressive priors in a disease mapping model: the Bayesian approach. *Biomedical Journal of Scientific and Technical Research*, 14.

Pickard, D. (1980). Unilateral Markov fields. *Advances in Applied Probability*, 12, 655-671.

Satria, R. and Castro, M. (2016). GIS tools for analyzing accidents and road design: A review. *Transportation Research Procedia*, 18, 242-247.

Scotto, M., Weiß, C. and Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: A review. *Statistical Modelling*, 15, 590-618.

Steutel, F. and Harn, K. V. (1979). Discrete analogues of self-decomposability and statibility. *The Annals of Probability*, 7, 3893-899.

Tevie, J., Bohara, A. and Valdez, R. (2014). Examination of the geographical variation in human west nile virus: a spatial filtering approach. *Epidemiology and Infection*, 142, 2522-2529.

Tjøstheim, D. (1983). Statistical spatial series modelling II: some further results on unilateral lattice processes. *Advances in Applied Probability*, 15, 562-584.

Tjøstheim, D. (1978a). A measure of association for spatial variables. *Biometrika*, 65, 109-114.

Tjøstheim, D. (1978b). Statistical spatial series modelling. *Advances in Applied Probability*, 10, 130-154.

Tobler, W. R. (1969). Geographical filters and their inverses. *Geographical Analysis*, 1, 234-253.

Valverde, J. and Jovanis, P. (2008). Analysis of road crash frequency with spatial models. *Transportation Research Record Journal of the Transportation Research Board*, 2061, 55-63.

Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, 8, 158-183.