Auto-models for spatially correlated counts

Nicole H. Augustin[1] and Jim McNicol[2]

[1] Zentrum für Datenanalyse und Modellbildung, Universität Freiburg

Eckerstrasse 1, D–79104 Freiburg im Breisgau

nicole@fdm.uni-freiburg.de

[2] Biomathematics and Statistics Scotland (BioSS), Scottish Crop Research Institute

Dundee DD2 5DA, UK

jim@bioss.sari.ac.uk

Abstract

In ecology data are frequently collected to observe the dynamics of vegetation. Often there are physical reasons for believing that the response of neighbours has a direct influence on the response at location i. For instance in an agricultural experiment competition might have a direct influence on yield. In terms of modelling such scenarios the family of auto-models or Markov random fields is an obvious choice. If the observed responses are counts, the auto-Poisson model can be used. There are different ways to formulate the auto-Poisson model, depending on the biological context. A drawback of this model is that for positive auto-correlation the likelihood of the auto-Poisson model is not available in closed form. We investigate how this restriction can be avoided by right truncating the distribution. We review different parameter estimation techniques which apply to auto-models in general and compare them in a simulation study. Results suggest that the method which is most easily implemented via standard statistics software, maximum pseudolikelihood, gives unbiased point estimates, but its variance estimates are biased. An alternative method, Monte Carlo maximum likelihood, works well but is computerintensive and not available in standard software. We illustrate the methodology and techniques for model checking with mite counts and seed count data from an agricultural experiment.

Keywords: auto-correlation, auto-Poisson model, competition, Monte Carlo maximum likelihood, Markov random fields, pseudolikelihood.

1 Introduction

When investigating the dynamics of vegetation and ecological processes the underlying questions are usually about: (a) The nature of local interaction among neighbouring plots for response variables, which can be none, competitive, antagonistic or co-operative. (b) The extent of spatial interaction, for example how big is the resource capture area of a plant, and how far do plants have to be apart in order not to compete with each other. (c) Finally in the types of relationship between spatially referenced response variables and covariates, for example, investigating whether the spatial distribution of plants or animals is related to spatial variation in environmental factors such as temperature. The last of these points can be dealt with using conventional regression methods, but ignoring the spatial correlation (the neighbour interaction) may lead to biased variance estimates and hence have an influence on deciding whether covariates have a significant effect on the response. If we want to describe ecological processes, modelling local interaction (autocorrelation) and underlying trend simultaneously is preferred as both features are often present.

Figure 1 shows simulated count data with different degrees of local autocorrelation. The model used for simulating the counts assumes different underlying processes including independence, trend along rows, positive autocorrelation and negative autocorrelation. It is not easy to establish different underlying processes by eye. In particular, the simulated data from the model assuming independence between squares does not seem much different to the simulated data from the model assuming positive autocorrelation. Thus there is clearly a need to estimate autocorrelation using modelling techniques. When dealing with survey or field experiment data (Figure 2), there are potentially many effects causing trends, e.g. soil gradients, altitude and treatments. These show up as positive autocorrelation, which makes it difficult to distinguish autocorrelation and trend in estimation. In contrast to the random-effects approaches (Besag et al., 1991; Breslow and Clayton, 1993; Wolpert and Ickstadt, 1998), we attempt to estimate autocorrelation directly as a fixed effect from the response. For this we examine the so-called auto-models first introduced by Besag (1972), to distinguish between the two components, trend and autocorrelation, with the aim of describing the underlying process. Our emphasis lies in estimation of the autocorrelation parameter and the specific neighbourhood structure it applies to, in contrast to epidemiological applications, where the emphasis is on predicting the spatial distribution.

Auto-models include models for counts, the auto-Poisson, auto-binomial and auto-negative binomial model. They also include models for continuous data such as the auto-normal model. The models are specified by the conditional probability of each observation given its set of neighbouring sites. The joint probability density, also called a Markov Random Field, is constructed using these conditional probabilities (Besag, 1974). These models incorporate a parameter for neighbour interaction (autocorrelation), caused by competitive or co-operative effects, which can influence global characteristics. A major drawback of this model is that it is only able to model negative autocorrelation. We investigate how this restriction can be avoided by right truncating the distribution. We give a short review on other existing methods for meodelling spatially correlated data in in section 2. In section 3 we derive and investigate properties of the truncated auto-Poisson and auto-negative binomial model. Since the auto-models incorporate dependence between responses, the likelihood is not always available in closed form and parameter estimation is not straightforward. In section 4 we describe how to simulate MRF data. In section 5 we describe and compare the different parameter estimation methods and show how Markov Chain Monte Carlo simulation can be used for model checking. A simulation study comparing the parameter estimation techniques is presented in section 6. Finally in section 7 we illustrate the methodology and model checking techniques with an analysis of mite and seed count data.

2 Review of existing methods

Many of the existing approaches for spatially correlated data have been developed for use with continuous data assuming a normal distribution. They include regression models based on simultaneous auto-regressive schemes (Whittle, 1954), moving average models, conditional auto-regressive (auto-normal) schemes (Cressie, 1991) and various forms of kriging (Cressie, 1991).

More recently generalised additive models (GAMs) (Hastie and Tibshirani, 1990) have been used to account for trend and correlation in normally distributed spatial data (Durbán et al., 1998; Gampe, 1998). GAMs are an extension of generalised linear models (GLMs) (McCullagh and Nelder, 1989). Both accommodate a variety of distributions for the response, but unlike GLMs, GAMs allow flexible nonlinear effects of the covariates on the response to be estimated from the data. The degree of smoothing performed by a smoother cannot be estimated directly using maximum likelihood, but it can be estimated using cross-validation (Wood, 2000). It is also possible to present smoothing splines as a mixed model, which allows estimation of smoothing parameters as a ratio of variance components via restricted maximum likelihood (Wang, 1998). Generalized linear mixed models (GLMMs) are models which are suited for spatially correlated counts, but they do not incorporate smoothers. Lin and Zhang (1999) have combined the two approaches for estimating cubic smoothing splines and spatial correlation parameters simultaneously via restricted maximum likelihood (REML). Fahrmeir and Lang (2001) fit Bayesian semiparameteric mixed models based on Markov random field priors.

The random-effects Poisson model first introduced by Besag et al. (1991) belongs to the class of GLMMs and has mainly been applied in epidemiology for disease mapping (Breslow and Clayton, 1993). This model can be used to account for autocorrelation, overdispersion and covariate effects for trend. In addition to the usual fixed linear covariate effects, the model contains an independent random location effect for heterogeneity and a spatially autocorrelated random effect. For the latter usually an auto-normal process (a Gaussian random field) with positive correlation over a defined neighbourhood is assumed. Similar to the above is a Bayesian hierarchical Gamma-Poisson model, which deals with positive and negative autocorrelation as well as dependence on other spatial covariates (Ickstadt and Wolpert, 1996; Wolpert and Ickstadt, 1998). For binary response variables there are applications where the auto-logistic model is used for direct modelling of autocorrelation. The wide range of applications covers meteorology (Hughes and Guttorp, 1994) and biology including genetics (Geyer and Thompson, 1992; Abel et al., 1993), plant epidemiology (Preisler, 1993; Preisler and Mitchell, 1993; Gumpertz et al., 1997), plant species mapping (Huffer and Wu, 1998; Wu and Huffer, 1997), and wildlife population mapping (Augustin et al., 1996, 1998). Other applications are similar to image restoration, where the auto-logistic model is used as a prior distribution for the random location effect. (Besag et al., 1991; Högmander and Møller, 1995; Heikkinen and Högmander, 1994; Hoeting et al., 2000).

Diggle et al. (1998) extend geostatistical methods to deal with non-normal data by formulating kriging as a generalised linear prediction model. The motivation of their work is in estimating spatial surfaces of non-normal variables, such as intensity of radioactivity and risk of infection. Another approach is to define a general covariance structure that accounts for autocorrelation. Gotway and Stroup (1997) use GLMs with a quasi likelihood, defined with a general variance-covariance matrix for spatially correlated data. They estimate the autocorrelation structure by fitting semi-variograms to residuals, after having fitted a GLM with trend only. As in the application of Diggle et al. (1998) their main concern is to obtain a better fit, rather than explaining the underlying process.

3 Description of models and parameter estimation

Assume that in an experiment or a survey we observe counts y_i at locations $i = 1, ..., n$. The counts could be, for example, the number of a particular species in a grid square or the number of fruit on a plant. If the distances between the locations are large, we can assume that the data are independent and that there is no neighbour interaction. We can fit a GLM and a Poisson or negative binomial distribution with a log link to these data:

$$\log(\boldsymbol{\mu}) \quad = \quad X\boldsymbol{\beta}$$

with $y_i \sim \text{Poi}(\mu_i)$ for which $Var(y_i) = \mu_i$ or negative binomial $y_i \sim \text{NegBin}(\mu_i, k)$, with $\mu_i = \frac{k}{\gamma i}$ and $Var(y_i) = \mu_i + \mu_i^2/k$, where k is the shape or aggregation parameter and the parameter $\gamma_i > 0$. The additional term μ_i^2/k in the variance function of the negative binomial distribution allows for more variation in the observations compared to the Poisson distribution. If counts have a Poisson distribution with variable rate, the rate itself having a gamma distribution, then the counts can be shown to have a negative binomial distribution (Johnson et al., 1992). As $k \mapsto \infty$, the negative binomial (NegBin) distribution tends towards a Poisson distribution. The parameter k can be estimated using an extra iterative step in the GLM procedure. For now we will concentrate on Poisson data. The matrix $\boldsymbol{X}$ is the design matrix containing explanatory variables and/or factors. With this model we can investigate relationships between spatially referenced response variables and explanatory variables, assuming the response $\boldsymbol{y}$ is independent.

If the counts are observed at adjacent locations, on a lattice structure D with sites i, $(i = 1, ..., n)$, we cannot always assume independence. In this case the auto-Poisson model can be used to describe the actual process underlying $\boldsymbol{Y}$ and to test for the type of interaction present. The auto-Poisson model is a Markov Random Field (MRF) originally developed in physics for magnetic fields. The classic version of a MRF is the first-order auto-logistic model developed for ferromagnetic applications, also called the Ising model (Ising, 1925). Auto-models are constructed by specifying the conditional probability of each observation given its set of neighbouring sites (Besag, 1974). The neighbourhood N_i could be for example the "first-order" type neighbourhood structure such as:

```
0 1 0
1 i 1
0 1 0
```

Sites in the lattice denoted by `1` are neighbours of site `i` and comprise the neighbourhood N_i. Note that N_i excludes site `i`. The joint probability density, also called a Markov Random Field, is constructed using the conditional probability of each observation.

For the Poisson and the negative binomial distributions the likelihood cannot be obtained directly by using the known conditional probabilities. But with the Markov random fields theorem (Besag, 1974) the likelihood function can be constructed in a straightforward manner. If the response y_i given N_i has any distribution of the exponential family, e.g. $\text{Poi}(\mu_i|y_j; j \in N_i)$, this theorem gives the so called Q(.)-function in terms of exponential family parameters. The Q-function is then used to define the likelihood: $l(\boldsymbol{\eta}) = Q(\boldsymbol{y}) - log\left(\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z}))\right)$. The parameterisation for members of the exponential family is

$$Pr(y_i|y_j; j \in N_i) = \exp\left(\ \theta_i(y_j; j \in N_i)B_i(y_i) + C_i(y_i) + D_i(y_j; j \in N_i)\ \right).$$

The parameter θ_i is called the natural parameter in the world of GLMs. It is the mean response transformed in some way and is what we are aiming to model, e.g. for the Poisson distribution $\theta_i = \log(\mu_i)$. The first derivative of the likelihood of any exponential distribution with respect to θ_i is equal to $\mu_i = E(y_i)$. The functions $B_i()$ and $C_i()$ have some specific form and $\theta_i()$ and $D_i()$ are functions of the neighbouring sites. Besag's theorem (Besag, 1974) assumes the above parameterisation and pairwise-only dependence between sites. As a consequence $Q(\boldsymbol{y})$ can be defined in terms of the exponential family parameters up to an additive constant (Cressie, 1991).

$$Q(y) = \sum_{i=1}^{n} (B_i(y_i) trend_i + C_i(y_i)) + \sum_{i=1}^{n-1} \sum_{j>i}^{n} B_i(y_i) B_j(y_j) \gamma_{ij}. \tag{1}$$

with i=1,...,n, $\gamma_{ij} = \gamma_{ji}, \gamma_{ii} = 0$ and $\gamma_{ik} = 0$ for $k \notin N_i$. In our case $trend_i = \boldsymbol{x}_i' \beta$. This gives the "recipe" to construct auto-models and makes it possible to formulate the model in terms of the natural parameter θ_i:

$$\theta_i(y_j; j \in N_i) = \boldsymbol{x}_i' \beta + \sum_{j \in N_i} \gamma_{ij} B_j(y_j). \tag{2}$$

This formulation is analogous to a GLM with a canonical link, except that here the observations themselves are explanatory variables. We call these *autocovariates*, with

$$autocovariate_i = \sum_{j \in N_i} B_j(y_j)$$

and often we assume that neighbourhood effects are the same in all directions, $\gamma = \gamma_{ij}$, so that

$$\sum_{j \in N_i} \gamma_{ij} B_j(y_j) = \gamma \ autocovariate_i.$$

If $Y_i \sim \text{Poi}(\mu_i | y_j; j \in N_i)$ for site i, the parameters in terms of the exponential family are $\theta_i(y_j; j \in N_i) = \log(\mu_i | y_j, j \in N_i)$, $B_i(y_i) = y_i$, $C_i(y_i) = -\log(y_i!)$ and $D_i(y_j; j \in N_i) = \mu_i(y_j; j \in N_i)$. Applying Besag's theorem, the auto-Poisson model formulated in terms of the natural parameter is thus

$$\log(\mu_i | y_j, j \in N_i) = \alpha_i + \sum_{j \in N_i} \gamma_{ij} y_j \tag{3}$$

with $\alpha_i = \boldsymbol{x}_i' \boldsymbol{\beta}$. Using equation 1 yields the $Q(.)$-function:

$$Q(\boldsymbol{y}) \ = \ \sum_{i=1}^{n} (\alpha_i y_i - \log(y_i!)) + \sum_{i=1}^{n-1} \sum_{j>i}^{n} \gamma_{ij} y_i y_j$$

where $\gamma_{ij} = \gamma_{ji}$, $\gamma_{ii} = 0$ and $\gamma_{ij} = 0$ for all $j \notin N_i$.

Unfortunately this model can describe negative correlation only, as the interaction parameter γ_{ij} is constrained to be negative (Besag, 1974): $\sum_{\boldsymbol{z} \in \zeta} exp(Q(\boldsymbol{z})) < \infty$ for all $\gamma_{ij} \leq 0$ only. We now propose a simple way to avoid having a normalising constant diverging to infinity: we truncate the distribution with a nominal truncation value r. The *right truncated Poisson* conditional probability mass function is then

$$Pr(Y_i = y_i | y_j; j \in N_i) = \frac{\mu_i^{y_i}}{y_i!} \left(\sum_{k=0}^{r} \frac{\mu_i^k}{k!} \right)^{-1}, \tag{4}$$

with $y_i = 0, 1, ..., r$ and the parameters in terms of the exponential family are the natural parameter $\theta_i(y_j; j \in N_i) = \log(\mu_i | y_j, j \in N_i)$, $B_i(y_i) = y_i$, $C_i(y_i) = -\log(y_i!)$ and $D_i(y_j; j \in N_i) = -\log\left(\sum_{k=0}^{r} \frac{\mu_i^k}{k!}\right)$. Hence we have

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n} (\alpha_i y_i - \log(y_i!)) + \sum_{i=1}^{n-1} \sum_{j>i}^{n} \gamma_{ij} y_i y_j. \tag{5}$$

The $Q(.)$-function of the truncated auto-Poisson model has the same expression as for the auto-Poisson model, but here $\sum_{z \in \zeta} exp(Q(\boldsymbol{z})) < \infty$ for all finite γ_{ij}.

An alternative to the truncated Poisson distribution is the Winsorized Poisson distribution that is used by Kaiser and Cressie (1997) for modelling Poisson variables with positive auto-correlation.

The normalising constant of the likelihood $\sum_{\boldsymbol{z} \in \zeta} exp(Q(\boldsymbol{z}))$ is taking the sum of the $Q(.)$-functions over the whole support of the distribution and as a consequence the likelihoods and parameter estimates could differ between models assuming a truncated or a Winsorized auto-Poisson model. If the value r is large relative to the mean of the distribution the difference between parameter estimates will be small; the shapes of the distributions do not differ. But if r is small, shapes differ more and it is best to use the most appropriate distribution. For instance if counts greater than r are censored, e.g. because like in our mite counts example more than r mites could not be counted, the Winsorized Poisson distribution is more appropriate.

For data which exhibit more variation than Poisson variables the negative binomial distribution is more appropriate. If there is need to account for auto-correlation, a

MRF can be specified with a NegBin distribution, yielding the auto-negative binomial model:

$$\log(\frac{1}{\beta_i + 1}) = \log(\frac{\mu_i}{k + \mu_i}) = \alpha_i + \sum_{j \in N_i} \gamma_{ij} y_j$$

where $\gamma_{ij} = \gamma_{ji}, \gamma_{ii} = 0$. Note that only the parameter β_i and hence μ_i is a function of the response in the neighbourhood of site i, not k. Following the steps as for the auto-Poisson model the Q-function can be derived using the Markov random field theorem.

Properties of the truncated auto-Poisson model

From now on we refer to the truncated auto-Poisson model of the following form as the **basic model**:

$$\log(\mu_i | y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} y_j. \tag{6}$$

for $y_j = 0, ..., r$ assuming $\gamma_{ij} = \gamma$ for all i and j and assuming a homogeneous process with $\alpha_i = \alpha$. Investigating the mean response $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ aids biological interpretation of parameters and determining whether the model is appropriate for explaining processes in vegetation dynamics. The auto-correlation parameter γ has a direct influence on $E(\bar{y}) = \mu_i$:

$$\mu_i \begin{cases} < exp(\alpha) & \text{if } \gamma < 0 \\ = exp(\alpha) & \text{if } \gamma = 0 \\ > exp(\alpha) & \text{if } \gamma > 0 \end{cases}$$

In the case where the observations in all neighbours are zero, $\sum_{j \in N_i} y_j = 0$:

$$\log(\mu_i | y_j; j \in N_i) = \alpha_i.$$

Thus $exp(\alpha_i)$ can be interpreted as the underlying *persistence rate*, given on the response scale. In Figure 3(a) the average $\bar{y}$ ($\bar{y}$ is the mean of a single realisation) from 500 simulated data sets, i.e. 500 realisations of a MRF as specified in equation 6, is plotted versus different values of γ. We explain how MRF's are simulated in the next section. A first-order neighbourhood is assumed and the *persistence rate* is $exp(\alpha) = 5$. For $\gamma < -1$ the average $\bar{y}$ is roughly equal to half the *persistence rate*. This is because for such strong negative correlation the pattern is similar to a chess board and half of the sites are surrounded by zeros (see also Figure 4 top graph). The dip in 3(a) correponds to an image similar to Figure 4 (top) but with some patches with y-values near zero.

There is no *lower limit* on γ due to the exponential link function, but as illustrated in Figure 3(a), any value of $\gamma < -1$ yields the same effect, also called a phase transition. The simulation results in Figure 3 also show that there is an *upper limit* for γ and $\gamma >$ *upper limit* leads to explosion. Modifying the truncated auto-Poisson model by subtracting the trend, $exp(\alpha_j)$, from the neighbouring y_j in equation 6 leads to a model with *no* unambiguous *persistence rate* (Figure 3b). Also centering the autocovariate does not help much (Figure 3c).

4 Simulating MRF data using the Gibbs sampler

An important prerequisite of parameter estimation of auto-models is the simulation of MRF data, which is also essential for fitted values and goodness-of-fit statistics. It is also required for the MCML method described in section 5. For a Gaussian MRF, also called an auto-normal model, data simulation is straightforward; it only involves simulating random white noise (see also Cressie, 1991, p. 202). For other distributions the joint density is generally not available in closed form and even when it is, it can be very computer intensive to simulate directly from it.

Given that our distribution of interest, the MRF defined by the auto-model, is not available in closed form and the fact that the full conditional densities for the response at locations i are available, the Gibbs sampler (Geman and Geman, 1984), a Markov chain Monte Carlo scheme (MCMC), is a natural choice for data simulation. In general the Gibbs sampler applies when the distribution of interest cannot be sampled from, but full conditional distributions are available. Gamerman (1997) gives a good summary on the general algorithm, which is based on successive generations from the full conditional distributions. We illustrate the algorithm for the case of the auto-Poisson model with

$$Y_i \sim \text{truncPoi}(\mu_i | y_j; j \in N_i),$$

$i = 1, ..., n$:

$$\log(\mu_i | y_j; j \in N_i) = \boldsymbol{x}_i' \boldsymbol{\beta} + \sum_{j \in N_i} \gamma_{ij} y_j. \tag{7}$$

The distribution of interest is the joint density $Pr(\boldsymbol{y})$ of the MRF with the above conditional specification. Thus the Gibbs sampler is implemented as follows:

1. Set the counter of the chain t to 1 and set initial values $\boldsymbol{y}^{(0)}$ by generating independent $\boldsymbol{Y}$ where $y_i^{(0)} \sim \text{Poi}(\mu_i)$ with $\log(\mu_i) = \boldsymbol{x}_i' \boldsymbol{\beta}$. The term $\boldsymbol{x}_i' \boldsymbol{\beta}$ is the part of the model in equation 7 describing the trend only.

2. Update $\boldsymbol{y}$ in the lattice successively by systematically scanning through the lattice or by picking locations at random and without replacement:

   $$y_i^{(t)} \sim \text{truncPoi}(\mu_i | y_j^{(t)}, y_j^{(t-1)}; j \in N_i)$$

   with μ_i as defined in equation 7. The neighbourhood N_i is calculated using the current realisation of the lattice, i.e. it can contain previously updated values $y_j^{(t)}$ and $y_j^{(t-1)}$, values which have not yet been updated.

3. Set $t = t + 1$ and return to step 2 until convergence is reached. Once convergence is reached, save realisations $\boldsymbol{y}^{(t)}$ at every lth step yielding the sample of size m $\boldsymbol{y}^{(c)}, \boldsymbol{y}^{(c+l)}, ..., \boldsymbol{y}^{(c+(m-1)l)}$, where c is the point of convergence.

To simulate truncated Poisson variables we use the Poisson distribution and reject values which are greater than the truncation value r. When consecutive realisations

$\boldsymbol{y}^{(t)}$ are highly correlated and each realisation is saved, a higher sample size is required than for the case of low correlation between iterations.

Simulation of auto-negative binomial MRFs follows the same procedure. We make use of the fact that the NegBin distribution can be derived as a mixture of Poisson distributions with rates distributed as Gamma$(\beta_i(y_j; j \in N_i), k)$ (Johnson et al., 1992). For sampling from $Y_i \sim$ trunc. NegBin$(\beta_i(y_j; j \in N_i), k)$, we first sample

$$\lambda_i \sim \text{Gamma}(\beta_i(y_j; j \in N_i), k),$$

for $\beta > 0, k > 0$ and then generate

$$y_i \sim \text{truncPoi}(\lambda_i | y_j; j \in N_i).$$

Convergence of the Gibbs sampler

We assessed the convergence of the Gibbs sampler by monitoring the following summary statistics of the likelihood:

$$\boldsymbol{t}(\mathbf{y}) = (\boldsymbol{t}_1, t_2)' = (\sum_i^n \boldsymbol{x}_i y_i, \frac{1}{2} \sum_i^n y_i autocovariate_i)',$$

where $\boldsymbol{t}_1$ corresponds to the summary statistic of the parameter $\boldsymbol{\beta}$ and t_2 corresponds to summary statistic of the parameter γ, assuming $\gamma = \gamma_{ij}$ and $autocovariate_i = \sum_{j \in N_i} y_j$ in the Q()-function of equation 5 or 7. In all cases the summary statistics are not highly correlated and do not remain on similar levels for any length of time, thus the Gibbs sampler moves around the target distribution very fast.

For a more formal assessment of convergence a method given in Gelman and Rubin (1992) and Gelman (1995) can be used. This method is based on comparing the between and within variance of parallel sequences of monitored summary statistics from a Markov chain. The idea is that with convergence the empirical distributions obtained separately from each sequence is approximately the same as the distribution obtained by mixing all sequences together. Before the sequences have converged the simulation collected within each single sequence will be much less variable than the simulation collected from all the sequences combined.

Utilising simulation for model checking

When we fit an auto-model to data and use estimated parameters to simulate a MRF as described above, the ergodic average $\bar{\boldsymbol{y}}$ of $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, ..., \boldsymbol{y}^{(m)}$ represents the approximate fitted values $\hat{\boldsymbol{\mu}}$ with components $\hat{\mu_i}$. This is justified by the following standard result:

$$\mu_i = E(y_i) = E(E(y_i|y_j; j \in N_i)) = \sum_{y_i \in \zeta} y_i Pr(y_i)$$

where $\zeta \equiv \{\boldsymbol{y} : Pr(\boldsymbol{y}) > 0\}$. The conditional expected value $E(y_i|y_j; j \in N_i)$ is estimated with the fitted value from the auto-model. For instance for the auto-Poisson model we use

$$\hat{E}(y_i|y_j, j \in N_i) = exp(\hat{\alpha}_i + \sum_{j \in N_i} \hat{\gamma}_{ij} y_j).$$

If the likelihood was available in terms of μ_i, deviance residuals and residual deviance could be used to assess goodness-of-fit. Instead we can plug the approximate fitted values $\bar{y}_i$ from ergodic averaging into the deviance function derived from likelihood functions assuming independence of observations. This likelihood corresponds to the pseudo likelihood function, which will be described below, but fitted values are not conditional on neighbourhood values. For example the deviance of the Poisson log-likelihood function is

$$D(y; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^{n} y_i log\left(\frac{y_i}{\mu_i}\right) - (y_i - \mu_i).$$

We can approximate the deviance by

$$D(y; \bar{\boldsymbol{y}}) = 2 \sum_{i=1}^{n} y_i log\left(\frac{y_i}{\bar{y}_i}\right) - (y_i - \bar{y}_i),$$

and calculate approximate deviance residuals

$$r_{iD} = sign(y_i - \bar{y}_i)\sqrt{2y_i log\left(\frac{y_i}{\bar{y}_i}\right) - (y_i - \bar{y}_i)}.$$

$D(y;\bar{\boldsymbol{y}})$ can be used as a goodness-of-fit statistic since it can be shown (McCullagh and Nelder, 1989) that $D(y;\hat{\boldsymbol{\mu}}) \sim \chi^2_{df}$, $df = n - p$. We assume that this is also the case for $D(y;\bar{\boldsymbol{y}})$. Also for comparing nested models the change of the approximated deviance $D(y;\bar{\boldsymbol{y}}_0) - D(y;\bar{\boldsymbol{y}}_A)$ can be used as usual to test for the significance of a covariate. Here $\bar{\boldsymbol{y}}_0$ corresponds to fitted values of the null model and $\bar{\boldsymbol{y}}_A$ corresponds to fitted values with the additional covariate.

Alternatively we can calculate the Pearson residual

$$r_{iP} = \frac{y_i - \hat{E}(y_i)}{\sqrt{(V(\hat{E}(y_i)))}} = \frac{y_i - \hat{E}(y_i)}{\sqrt{(\hat{E}(y_i^2) - \hat{E}(y_i)^2)}}$$

which yields the Pearson χ^2 goodness-of-fit statistic (McCullagh and Nelder, 1989)

$$\sum_{i=1}^{n} r_{iP}^2 = X^2,$$

$X^2 \sim \chi^2_{df}$, $df = n - p$, where n is the number of observations and p the number of parameters. We replace $\hat{\mu}_i = \hat{E}(y_i)$ by its ergodic average $\bar{y}_i$.

By replacing fitted values $\hat{\mu}_i$ with their ergodic averages $\bar{y}_i$, we can obtain the ingredients needed for classical model checking and goodness-of-fit statistics. Further work is needed to investigate the distributional properties of the deviance and Pearson X^2-statistic.

5 Parameter estimation for auto-models

Parameter estimation for auto-models using full **maximum likelihood (ML)** is only possible if $\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z}))$ is available in closed form. Then the log-likelihood has the form:

$$l(\boldsymbol{\eta}) = Q(\boldsymbol{y}) - log\left(\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z}))\right). \quad (8)$$

where $\boldsymbol{\eta}$ is the parameter vector and $\boldsymbol{y} \in \zeta$, $\zeta \equiv \{\boldsymbol{y} : Pr(\boldsymbol{y}) > 0\}$. For the auto-logistic model the likelihood is available in closed form, but for most data sets evaluating the normalising constant is computationally too intensive. For the truncated and Winsorized auto-Poisson model the normalising constant can also be evaluated: Kaiser and Cressie (1997) derived full maximum likelihood estimates for a Winsorized auto-Poisson model. As for the auto-logistic model the procedure is very computer intensive. Kaiser and Cressie (1997) use a simulation example comprising six locations to test the influence of different truncation points on parameter estimates. Deriving the normalising constant for a data set with moderate size full maximum likelihood estimation becomes infeasible and other possibilities for parameter estimation need to be used. Other possibilities for parameter estimation are given below.

The **pseudo likelihood (PL)** function (Besag, 1975) is the product of the conditional probabilities, ignoring non-independence of observations. For example for the auto-Poisson model the pseudo log-likelihood has the form:

$$
\begin{aligned}
l_p(\boldsymbol{\eta}) &= \sum_{i=1}^{n} log(Pr(y_i|y_j; j \in N_i)) \\
&= \sum_{i=1}^{n} log((\mu_i|y_j; j \in N_i) + y_i log(\mu_i|y_j; j \in N_i) - log(y_i!))
\end{aligned}
$$

By formulating the auto-model as a GLM with the canonical link, maximum pseudo likelihood (MPL) estimation can be performed in any GLM package using iteratively re-weighted least squares. Because maximum pseudo likelihood assumes observations are independent, standard errors from iteratively re-weighted least squares are likely to be biased when strong autocorrelation is present. The parametric bootstrap can be applied to obtain standard errors, in which case the data are simulated using the Gibbs sampler.

Besag (1974) introduced **coding**, where the lattice is subdivided into two or more disjoint sub-lattices, so that the sites of the sub-lattices are mutually independent. Coding estimators are obtained by maximising the conditional likelihood of one sub-lattice given the other. This yields two or more possible coding estimates that can be combined in some way. For example if correlation is assumed only among the four nearest neighbours, there are two sub-lattices D_0 and D_1, whose elements are mutually independent. Then the conditional log-likelihoods are: $l_{c,D_0}(\eta) = \sum_{i \in D_o} log(Pr(y_i|y_j : j \in N_i))$ and $l_{c,D_1}(\eta) = \sum_{i \in D_1} log(Pr(y_i|y_j : j \in N_i))$. Estimates from the two conditional likelihoods are combined.

Monte Carlo maximum likelihood (MCML) was introduced by Geyer and Thompson (1992). It is very similar to Monte Carlo Newton-Raphson introduced by Penttinen (1984). It can be applied to any model for dependent data derived from the exponential family such as MRFs. Other examples include spatial models for point processes, models for Markov graphs and conditional maximum likelihood. We illustrate the method with the right truncated auto-Poisson model from equation 4. We rewrite $Q(.)$ and $l(\boldsymbol{\eta})$ from equation 8 as

$$Q(\boldsymbol{y}) = \boldsymbol{t}(\boldsymbol{y})'\boldsymbol{\eta} \quad = \quad \sum_{i=1}^{n}(\alpha_i y_i - log(y_i!)) + \sum_{i=1}^{n-1}\sum_{j>i}^{n} \gamma_{ij} y_i y_j.$$

Considering the log-likelihood as a function of the parameters $l(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ the term $\sum_{i=1}^{n} -log(y_i!)$ is a constant and can be omitted, so that (by redefinition of $\boldsymbol{t}(\boldsymbol{y})$).

$$\boldsymbol{t}(\boldsymbol{y})'\boldsymbol{\eta} \quad = \quad \sum_{i=1}^{n-1} \alpha_i y_i + \sum_{i=1}^{n-1}\sum_{j>i}^{n} \gamma_{ij} y_i y_j,$$

with vectors $\boldsymbol{t}(\boldsymbol{y}) = (y_1, ..., y_n, y_1 y_2, ..., y_{n-1} y_n)'$ and $\boldsymbol{\eta} = (\alpha_1, ..., \alpha_n, \gamma_{12}, ..., \gamma_{n-1,n})'$. Then the log-likelihood takes the form

$$l_y(\boldsymbol{\eta}) \quad = \quad log\left(\frac{exp(\boldsymbol{t}(\boldsymbol{y})'\boldsymbol{\eta})}{c(\boldsymbol{\eta})}\right)$$

with the normalising constant

$$c(\boldsymbol{\eta}) = \sum_{\boldsymbol{z} \in \zeta} exp(\boldsymbol{t}(\boldsymbol{z})'\boldsymbol{\eta}).$$

Now a constant term $log(c(\boldsymbol{\psi}))$ is added to the likelihood. The term $c(\boldsymbol{\psi})$ has the same form as the normalising constant $c(\boldsymbol{\eta})$, vectors $\boldsymbol{\psi}$ and $\boldsymbol{\eta}$ are of the same length and the values of $\boldsymbol{\psi}$ are constant and close to $\boldsymbol{\eta}$. The log-likelihood is then

$$\begin{aligned} l_y(\boldsymbol{\eta}) &= log\left(\frac{exp(\boldsymbol{t}(\boldsymbol{y})'\boldsymbol{\eta})}{c(\boldsymbol{\eta})}\right) + log(c(\boldsymbol{\psi})) \\ &= \boldsymbol{t}(\boldsymbol{y})'\boldsymbol{\eta} - log\left\{\frac{c(\boldsymbol{\eta})}{c(\boldsymbol{\psi})}\right\} \end{aligned}$$

It can be shown (Geyer and Thompson, 1992) that the ratio $\frac{c(\boldsymbol{\eta})}{c(\boldsymbol{\psi})}$ can be expressed as the expected value $E_{\boldsymbol{\psi}}(exp(\boldsymbol{t}(\boldsymbol{y})'(\boldsymbol{\eta}-\boldsymbol{\psi}))$, which can be approximated by its ergodic average

$$E_{\boldsymbol{\psi}}(exp(\boldsymbol{t}(\boldsymbol{y})'(\boldsymbol{\eta}-\boldsymbol{\psi})) \approx \frac{\bar{c}(\boldsymbol{\eta})}{\bar{c}(\boldsymbol{\psi})} = \frac{1}{m}\sum_{i=1}^{m} exp(\boldsymbol{t}(\boldsymbol{y}^{(i)})'(\boldsymbol{\eta}-\boldsymbol{\psi}))$$

with $\boldsymbol{t}(\boldsymbol{y}^{(i)})$ simulated from an ergodic Markov chain $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, ..., \boldsymbol{y}^{(m)}$ with equilibrium $P_{\boldsymbol{\psi}}(\boldsymbol{y})$, the joint density of the MRF with the auto-Poisson model defined by parameters $\boldsymbol{\psi}$. The approximate log-likelihood is then

$$l_{\boldsymbol{y},m}(\eta) = \boldsymbol{t}(\boldsymbol{y})'\boldsymbol{\eta} - log\left(\frac{1}{m}\sum_{i=1}^{m} exp(\boldsymbol{t}(\boldsymbol{y}^{(i)})'(\boldsymbol{\eta}-\boldsymbol{\psi}))\right)$$

Note that $\boldsymbol{t}(\boldsymbol{y}), \boldsymbol{\eta}$ and $\boldsymbol{\psi}$ are all vectors of length p, the number of parameters, and the term $exp\left(\boldsymbol{t}(\boldsymbol{y}^{(i)})'(\boldsymbol{\eta}-\boldsymbol{\psi})\right)$ is a scalar. Maximising this approximate likelihood yields the Monte Carlo maximum likelihood estimator (MCMLE) $\hat{\boldsymbol{\eta}}_{y,m}$. The approximate likelihood $l_{\boldsymbol{y},m}(\boldsymbol{\eta})$ is only a good approximation to $l_{\boldsymbol{y}}(\boldsymbol{\eta})$ near $\boldsymbol{\psi}$, so that iterative constrained maximisation is needed. The estimated standard error of the MLE can be derived as the square root of the diagonal elements of the inverse observed Fisher information of $l_{\boldsymbol{y},m}(\boldsymbol{\eta})$, which is the negative Hessian matrix $-H(\boldsymbol{\eta})$.

Figure 5 gives a schematic overview of how MCML is performed in practice. For the initial approximation of the likelihood, $\boldsymbol{\psi}_r$ is set to parameter estimates from another method such as MPL, coding or ML of a model assuming independence, i.e. without the parameter γ. Then auto-Poisson data with parameters $\boldsymbol{\psi}_r$ are simulated using the Gibbs sampler and using these simulated data in the normalising constant the approximated likelihood can be derived and maximised. The maximisation is constrained to be within an upper and lower limit around $\boldsymbol{\psi}_r$. If the obtained estimates

$\hat{\boldsymbol{\eta}}_r$ are close to $\boldsymbol{\psi}$ the approximation should be good enough and the $\hat{\boldsymbol{\eta}}_r$ are the MCML estimates. If they are not, we start again simulating data with $\boldsymbol{\psi}_{r+1} = \hat{\boldsymbol{\eta}}_r$ and so forth. We start from the MPL estimates in most examples. These are usually quite close to the MCML estimates, and convergence can be rapidly achieved.

The **Monte Carlo Newton-Raphson method** proposed by Penttinen (1984) is similar to Geyer's (Geyer and Thompson, 1992) MCML method. It has recently been applied to a Winsorised auto-Poisson model by Lee and Kaiser (1997). Other descriptions of the algorithm are given in Heikkinen and Penttinen (1999) or Geyer (1999). Rather than approximating the likelihood as in MCML, here only the gradient and Hessian matrix are approximated. For the update,

$$\boldsymbol{\eta}_{r+1} = \boldsymbol{\eta} - (H(\boldsymbol{\eta}_r))^{-1} \frac{\delta l}{\delta \boldsymbol{\eta}_r},$$

where

$$\frac{\delta l}{\delta \boldsymbol{\eta}} = \boldsymbol{t}(\boldsymbol{y}) - E_{\boldsymbol{\eta}}(\boldsymbol{t}(\boldsymbol{y}))$$

and

$$H(\boldsymbol{\eta}) = -Var_{\boldsymbol{\eta}}(\boldsymbol{t}(\boldsymbol{y})).$$

In each update of the Newton-Raphson algorithm a sample $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, ..., \boldsymbol{y}^{(n)}$ from a Markov Chain with equilibrium $P_{\boldsymbol{\eta}_r}(\boldsymbol{y})$ is drawn. The gradient $\frac{\delta l}{\delta \boldsymbol{\eta}}$ is replaced by $\boldsymbol{t}(\boldsymbol{y}) - \boldsymbol{t}(\bar{\boldsymbol{y}})$, where $\boldsymbol{t}(\bar{\boldsymbol{y}})$ is the ergodic average of $\boldsymbol{t}(\boldsymbol{y})$ and $H(\boldsymbol{\eta})$ is replaced by the negative sample covariance matrix.

Younes (1988) first introduced a **stochastic gradient algorithm** based on the Gibbs sampler to obtain maximum likelihood estimates. Geyer (1999) proposes using this method as a starting point for MCML. The idea is to run a Markov chain with non-stationary transition probabilities, adjusting the parameter $\boldsymbol{\eta}$ in each iteration to move towards the MLE. The method does not supply estimates for standard errors, which is why we do not pursue this method further here.

6 Comparison of different parameter estimation methods: A simulation study

The three parameter estimation methods, pseudo likelihood (MPL), coding (CODE1, CODE2) and Monte Carlo maximum likelihood (MCML) were compared in a simulation study. We simulate $Y_i \sim \text{truncPoi}(\mu_i|y_j; j \in N_i)$, $i = 1, ..., n$ defined by the model:

$$\log(\mu_i|y_j; j \in N_i) = \alpha + \beta x_i + \gamma \frac{1}{n_i} \sum_{j \in N_i} y_j. \tag{9}$$

where α is the intercept, x_i is a covariate for row effect with $x_i = row_i - \frac{1}{n}\sum_{i=1}^{n} row_i$, with slope parameter β. The parameter γ is the interaction parameter defining spatial autocorrelation, N_i are the four nearest neighbours and $n_i = 4$, except for edge sites. Respective parameter values for *studies 1* to *3* are given in Table 1, and those for *study 4* in 2. In *studies 1* to *3* the parameter β was zero, so that no trend was present. A truncation value of $r = 100$ for the MCML method was assumed, except for the second simulation study with $r = 20$. For the MCML method the initial value for $\boldsymbol{\eta}_{init}$ was taken from the model fitted assuming independence for those studies where the model was without auto-correlation (study 1 and 4). For studies 2 and 3, where the model included auto-correlation, the starting value was taken from the MPL estimates. This choice of $\boldsymbol{\eta}_{init}$ was based on the speed of convergence.

When comparing the mean asymptotic standard errors with empirical standard errors in Table 1 and 2, it appears that MPL underestimates standard errors in the simulation studies which incorporated zero or positive auto-correlation. In studies 1, 2 and 4 the mean asymptotic standard errors are substantially lower than empirical standard errors for the MPL method. This is also the case for the MCML method, probably caused by the extra component of Monte Carlo variation, measured by the Monte Carlo standard error (mcse), but differences there are smaller. Unfortunately the mcse cannot be used to inflate the variance correctly, because the covariance $Cov((\hat{\boldsymbol{\eta}}_{y,m} - \hat{\boldsymbol{\eta}}), (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}))$ is not known. In simulation *study 3*, where negative auto-correlation is incorporated, asymptotic standard errors from MPL are higher than their empirical counterparts.

The bias in parameter estimates in all methods appears to be negligible, given the associated standard errors. It is surprising to see that the coding method gives the lowest bias only in *study 2* for which $\gamma = 0.05$. This is probably because the coding estimates are derived from only half of the sample size. On average MPL gives the lowest bias in this limited investigation. In study 3, where $\gamma = -0.5$, MCML gives the largest bias, underestimating α and overestimating γ.

7 Examples

7.1 Mite count toy example

The mite data are used by Lee and Kaiser (1997) to illustrate their proposed Monte Carlo Newton-Raphson (MCNR) estimation method for the Winsorized auto-Poisson model. The data are taken from Hairston et al. (1971) who examined distribution of soil arthropods. Herbivorous mite counts were observed in 1965 on an 8 x 8 inch square of an abandoned field with continuous grass cover. A square was subdivided into 64 one inch cubes. These were labeled and taken to the lab, where the mites in each cube were counted. The data are shown in Figure 8. Similarly to Lee and Kaiser (1997) who fit a Winsorized auto-Poisson model, we fit a truncated auto-Poisson model to these data, using a truncation value $r = 7$ and assuming a first-order neighourhood interaction:

$$\log(\mu_i | y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} y_j \tag{10}$$

Figures 7 (a) and (b) show the sequence of $\hat{\alpha}$ and $\hat{\gamma}$ for the MCML procedure, together with the MPL estimate and that of a model assuming independence, that is equation 10 without the term $\gamma \sum_{j \in N_i} y_j$. The MCML method clearly converges. The MCML result is also very similar to Lee and Kaiser's result (Table 8).

The standard errors from MCML and MCNR are almost identical and if a 95% confidence interval for $\hat{\gamma}$ was constructed using the standard error (assuming standard maximum likelihood theory) then it would only just include zero. This leads us to the conclusion that both the truncated auto-Poisson model and the model assuming independence might be adequate. Standard errors from MPL are smaller, but MPL possibly underestimates the variance when positive auto-correlation is present. The cloud of simulated summary statistics $t_1^{(t)} = \sum y_i^{(t)}$ and $t_1^{(t)} = \sum y_i^{(t)} \sum_{j \in N_i} y_j^{(t)}$, calculated from realizations of a simulated ergodic Markov chain using MCML, MPL for the model in equation 10 and ML (independence) estimates (Figure 7(c) - (d)) all contain the observed summary statistic. This demonstrates that all of the models are plausible, but the independence model is marginal, because the observed summary statistic is at the edge of the cloud. Also the Pearson statistic of goodness-of-fit indicates that the fit of the independent models is not as good as for the MCML and the MPL model.

In contrast to that, investigation of correlograms of Pearson residuals in Figure 8 showed that the positive correlation of 0.15 present in the first lag does not differ between the auto-Poisson and independence model.

Long Ashton seed counts

The data are from a field experiment with a randomised block design which was conducted at Long Ashton Research Station. There were four blocks (16 columns and 5 rows), two crops (WW = winter wheat, SW = spring wheat), two cultivation types (P = plough, T = tine) and five weed groups. A split-plot design was used, i.e. cultivation type and crop were only randomised among columns. The plot size was 3 x 3 m, and plots were 5 m apart in the north-south direction and 2 m apart in the east-west direction. Here the counts of a perennial weed species *Poa* spp. (Figure 9) are analysed. Seed counts refer to the total number of seeds found in six soil cores of volume 0.2 l, taken at 0 to 10 cm depth, which were collected in four consecutive years. Here only the first year is analysed. The aim of the experiment was to observe persistence of seeds in the soil of the five groups of weed species sown at the start of the experiment, and to examine the persistence of *Poa* seeds which were already present before the experiment took place.

Figures 9 show the observed seed counts. Given the fact that counts are aggregated from six soil cores per plot, there is not much chance that "true" spatial autocorrelation, such as competition, can be observed. Therefore the analysis strategy is initially to fit a model with block and treatment effects, then residuals are checked for remaining autocorrelation. The experiment design is such that effects of cultivation type and crop are confounded with columns. Since we are interested in the effects of the treatments rather than prediction, we fit block and treatment effects rather than column effects.

Poa seed counts were low with a maximum of 9. We consider a GLM assuming either a Poisson or a negative binomial distribution. Let y_{it} be the seed count at site i. The maximal model considered is:

$$\begin{aligned}\log(\mu_i) &= \mathbf{x}_i'\boldsymbol{\beta} \\ &= block_i + weed_i * crop_i * cult_i\end{aligned}$$

with $y_i \sim \text{Poi}(\mu_i)$ or $y_i \sim \text{NegBin}(\mu_i, k)$, With both distributions a log-link is used. The best fit for *Poa* seed was obtained for the model

$$\log(\mu_i) = block_i + weed_i + crop_i + cult_i + weed_i.cult_i, \tag{11}$$

for both the NegBin and the Poisson distribution. The NegBin model has an estimated aggregation parameter of $\hat{k} = 3.32$ (s.e.$=$ 2.07) and the residual deviance is lower than that from the Poisson model (Table 4). The column effect is confounded with crop and cultivation type and not fitted.

Deviance residual plots (not shown) show that there is little difference between the Poisson model and the NegBin model. After adjusting for overdispersion in the Poisson case, estimates and standard errors for the fixed effects are very similar. Spatial correlation in the Pearson residuals using correlograms in the north/south and east/west directions shows that in both models there is autocorrelation remaining in the north/south directions (Figure 10), which is slightly weaker for the NegBin model. The autocorrelation within a column could be caused by the tractor trailing along the columns. We can account for the autocorrelation by incorporating a specific term for autocorrelation in the north/south direction using the truncated auto-Poisson model or the truncated auto-negative binomial model:

$$\log(\mu_i | y_j : j \neq i) = \mathbf{x}_i'\beta + \frac{1}{n_i}\gamma_{ij} \sum_{j \in N_i} y_j;$$

with $y_i \sim \text{Poi}(\mu_i | y_j; j \in N_i)$ or $y_i \sim \text{NegBin}(\mu_i | y_j; j \in N_i; k)$, n_i is the number of neighbours, N_i are the nearest four neighbours, except for edge plots, and $autocov_i = \frac{1}{n_i}\sum_{j \in N_i} y_j$. We define $\gamma_{ij} = ns$ for y_j which are north/south neighbours and $\gamma_{ij} = ew$ for y_j which are east/west neighbours. Parameter estimation for the auto-models can be carried out using pseudo likelihood, coding or the MCML method described in Section 5. When fitting this model using the MPL, only the north/south neighbourhood interaction is significant. Although standard GLM software provides the deviance of the pseudo likelihood when fitting the auto-Poisson model using the MPL method, it is not clear whether the deviance is approximately χ^2-distributed.

In addition to MPL, the MCML method was applied to obtain estimates for the auto-Poisson model. For MCML, the vector of summary statistics is

$$t(y) = (\sum y_i x_{i,1}, ..., \sum y_i x_{i,14}, \frac{1}{2}\sum y_i autocov_i)',$$

where

$$autocov_i = \frac{1}{n_i}\sum_{j \in N_i} y_j.$$

We have 14 columns in the design matrix of the independence model from equation 11. The parameter vector is

$$\boldsymbol{\eta} = (\beta_1, ..., \beta_{14}, ns)'.$$

One problem in this example is that the estimate for ns is at the upper boundary of the parameter space. This means that the MCML method cannot work very well, because simulation of data needed for the normalizing constant is restricted to values below or equal to the upper boundary. The MCML method breaks down, because estimates for γ reach the boundary and the counts simulated by the Gibbs sampler are at the truncation value. The other problem is that treatment, block and autocorrelation effects cannot be distinguished. When keeping the treatment and block parameter estimates from the Poisson model fixed and then estimating γ with MCML, the estimate converges. Using the converged $\hat{\gamma}$ we do one final run of MCML where all parameters are estimated simultaneously, and the resulting $\hat{\gamma} = 0.034$ ($s.e. = 0.056$) is our final estimate. It is only a quarter of the MPL estimate $\hat{\gamma} = 0.169$ ($s.e. = 0.072$) and a 95% confidence interval ($\gamma + / - 1.96 * s.e.(\gamma)$) for the MCML estimate would include 0, indicating that the autocorrelation is not significant. Investigating the correlograms in the north/south (Figure 10a)) and east/west (Figure 10b) directions of Pearson residuals from the three models (Figure 10), using the MCML estimates for the auto-Poisson model, shows no big difference between models. This is confirmed by a formal test of spatial correlation on the Pearson residuals using Moran's coefficient of correlation (Moran, 1950), which does not show significant positive correlation for any of the models when based on a first order neighbourhood (Table 4). Calculating Moran's coefficient of correlation based on a north/south neighbourhood, residuals from none of the models are significantly correlated, based on a 5% significance level. So in addition to the confidence intervals for the autocorrelation parameter ns from MCML estimates, correlograms and Moran's coefficient indicate that autocorrelation is not significant. In contrast to this is the estimate of ns we obtain from MPL, for which confidence intervals indicate a significant effect. Calculation of Pearson residuals for the auto-models requires simulating data using the Gibbs sampler, as described in Section 4. With MPL estimates this could not be achieved, counts simulated by the Gibbs sampler are at the truncation value. It has been quoted in the literature that when estimates are near the critical value where phase transition occurs, MPL estimates are more variable than MCML estimates (Geyer, 1991). All these encountered problems add to the evidence that the model with autocorrelation as it stands is not appropriate. Figure 10b) shows that there is negative correlation in east/west direction at a lag of 15 m. Including an additional autocovariate for the counts of the third plot in east and west direction does not improve the model fit.

8 Discussion

When choosing a spatial model there has to be some understanding of the underlying biological process. The basic auto-Poisson model for counts features an underlying persistence rate for locations with zero counts in their neighbourhood. If the counts are from vegetation, such an assumption is reasonable. But when the counts are from animals it is possibly inappropriate. An example for such a case is a population of mammals on an island where migration can be excluded.

Modelling auto-correlation for stationary processes is relatively straightforward. For a heterogeneous process things become more complicated; in particular positive auto-correlation cannot be distinguished from trend. As illustrated by the weed seed bank example, what appears as positive auto-correlation is often just caused by an underlying trend. Often the information to model this trend adequately is not available, because data collection on soil, habitat and physical characteristics would be too expensive. Generalized additive models (GAMs) provide a flexible tool to fit two-dimensional surfaces, which should be fairly good for estimating the trend. The combination of

GAMs and models with an explicit auto-correlation structure, such as auto-models, deserves further investigation.

If estimation of the auto-correlation parameters rather than prediction is the main aim, correlated parameter estimates in the auto-Poisson model is a problem. There are a few examples in the literature which exemplify that correlated parameter estimates are a common problem in spatial statistics. Kaiser and Cressie (1997) analyse simulated data using the Winsorized auto-Poisson model using maximum likelihood and obtain a correlation of -0.939, as estimated from the Fisher information matrix, between the estimates of intercept and autocorrelation parameter. There is an example in Geyer (1999) for a spatial model fitted to point process data where the correlation between estimates of trend and autocorrelation is -0.983. Because of such high correlation between estimates of the trend parameter α_i and the auto-correlation parameter γ a different method for constructing confidence intervals should be considered. If one of the parameters can be considered as a nuisance parameter, profile likelihood can be used. For instance consider α_i as a nuisance parameter. Obtain $\hat{\alpha}_i$ while keeping γ at a constant, then use the profile likelihood $l_{z,m}(\hat{\alpha}_i, \gamma)$ for constructing likelihood intervals for γ. If the data shows trend, $\boldsymbol{\beta}$ from the term $x_i'\boldsymbol{\beta}$ could be treated as nuisance parameters or for a homogeneous process the model could be fitted without an intercept. This and other possibilities for solving the unidentifiablity problem deserves further investigation.

The simulation study showed that MCML works well for estimating the parameters and standard errors. Also standard errors are asymptotically normal. But MCML is quite computer intensive and a lot more complicated than the MPL method. Problems we encountered with MCML turned out to be caused by a constraint on the upper boundary for the auto-correlation parameter γ in the auto-Poisson model. We have also shown that MPL gives unbiased estimates, but that its asymptotic standard errors are possibly biased. For the auto-logistic model of both MPL and MCML, estimates are consistent, but near the critical value of the autocorrelation parameter γ where phase transition starts, the variability of PLE estimates can be much higher than MCML estimates (Geyer, 1991). MCML has also been compared with MPL for the auto-logistic model by Wu and Huffer (1997), who also conclude that the MCML method yields more accurate results in terms of standard errors and mean square error than the MPL method, but MCML estimates are more biased than MPL estimates.

This means that, if we are interested in estimating parameters and we expect strong autocorrelation, MPL is not advisable. Also MPL tends to overestimate standard errors when there is negative correlation and underestimate standard errors when correlation is positive in the presence of strong correlation. Standard errors of the coding method always exceed those of both MPL and MCML. Although we only assume models with a first-order neighbourhood structure that lead to two sub-lattices in the coding method, this results in a substantial reduction of sample size ($\frac{n}{2}$) leading to the increased standard errors. With higher order neighbourhoods the standard errors will increase even more. Given that generally the neighbourhood structure is unknown and different neighourhood sizes need to be tested, coding is not a pragmatic option.

The truncated auto-Poisson model has the problem that the usual 95% confidence intervals could exceed the upper boundary of the auto-correlation parameter γ. In study 2 the simulated data had only mild positive auto-correlation ($\gamma = 0.05$), so that data simulation for the MCML method was still possible. It is quite likely that for higher values of γ, close to the upper boundary, asymptotic normality does not hold and confidence intervals will not be symmetric. In such cases alternative ways for quantification of precision should be found. One possibility is likelihood intervals.

The mite example showed that the MCML parameter estimation method works well for moderate levels of positive auto-correlation. The model checking procedures gave different answers as to which model is appropriate. This example was mainly included for illustrative purposes and further investigation of the appropriate neighbourhood structure over which autocorrelation applies is needed. We have seen in the seed bank example that the auto-Poisson models can be useful to improve model fit, even if the auto-correlation observed is just unmodelled trend. At the same time the example illustrates the difficulty of parameter estimation in cases when positive autocorrelation is present, typically caused by an unmodelled trend rather than true autocorrelation.

The auto-Poisson model has potential for modelling processes in vegetation dynamics. It is very flexible and with slight modifications, such as adding extra terms for trend, can accommodate a variety of processes. It can be applied even in cases with weak positive auto-correlation caused by unmodelled trend and can improve precision. When strong positive autocorrelation is present, parameter estimation becomes difficult, because estimates will be close to the upper boundary of the parameter space. Often strong autocorrelation indicates unmodelled trend, so in theory by adequately modelling the trend, boundary problems could be avoided. If data exhibit more variation than Poisson, the auto-negative binomial model can be used, but its properties do not allow positive auto-correlation to be modelled. As for the auto-Poisson case the model is only valid for negative auto-correlation and by truncating the distribution at the right end this can be circumvented. Unfortunately the restriction $\beta > 0$ leads to a very low upper limit for γ. This makes the (truncated) auto-negative model for positive correlation extremely difficult to work with, and for highly overdispersed count data with auto-correlation, other methods such as gamma-Poisson random fields (Wolpert and Ickstadt, 1998; Brix, 1998) or Gaussian-Poisson random fields (Besag et al., 1991; Breslow and Clayton, 1993) are recommended.

Acknowledgements

The first author was funded by the Scottish Executive Rural Affairs Department, research grant FF81695.

References

Abel, L., Golmard, J. L., and Mallet, A. (1993). An autologistic model for the genetic-analysis of familial binary data. *American Journal of Human Genetics*, 53(4):894–907.

Augustin, N. H., Mugglestone, M. A., and Buckland, S. T. (1996). An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33:339–347.

Augustin, N. H., Mugglestone, M. A., and Buckland, S. T. (1998). The role of simulation in modelling spatially correlated data. *Environmetrics*, 9:175–196.

Besag, J. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society, B*, 34:75–83.

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

Besag, J. (1975). Statistical analysis of non-lattice data. *Statistician*, 24:179–195.

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

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25.

Brix, A. (1998). *Spatial and Spatio-temporal Models for Weed Abundance*. PhD thesis, Royal Veterinary and Agricultural University Copenhagen, Denmark.

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

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society, C - Applied Statistics*, 47:299–326.

Durbán, M. L., Hackett, C. A., and Currie, I. (1998). Blocks, trend and interference in field trials. In Marx, B. and Friedl, H., editors, *Statistical Modeling. Proceedings of the 14th International Workshop on Statistical Modeling.*, Graz, Austria.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on markov random field priors. *Applied Statistics*, 50:201–220.

Gamerman, D. (1997). *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference.* Chapman & Hall, London.

Gampe, J. (1998). Trend or correlation? - attributing smoothness in nonparametric regression. In Marx, B. and Friedl, H., editors, *Statistical Modeling. Proceedings of the 14th International Workshop on Statistical Modeling*, pages 216–221, New Orleans, Louisiana.

Gelman, A. (1995). Inference and monitoring convergence. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 131–161. Chapman & Hall, London.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7:457–511.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *I. E. E. E. Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. In *Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pages 156–163.

Geyer, C. J. (1999). Likelihood inference for spatial point processess. In Barndorff-Nielsen, O. E. and Kendall, W. S., editors, *Current Trends in Stochastic Geometry and Applications (to appear)*. Chapman and Hall, London.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, B*, 54:657–699.

Gotway, C. A. and Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 2:157–178.

Gumpertz, M. L., Graham, J. M., and Ristaino, J. B. (1997). Autologistic model of spatial pattern of phytophthora epidemic in bell pepper. *Journal of Agricultural, Biological, and Environmental Statistics*, 2:131–156.

Hairston, N. G., Hill, R. W., and Ritte, U. (1971). The interpretation of aggregation patterns. In *Statistical Ecology Volume 1: Spatial Patterns and Statistical Distributions.* The Pennsylvania State University Press, University Park, Pennsylvania.

Hastie, T. and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Heikkinen, J. and Högmander, H. (1994). Fully Bayesian approach to image restoration with an application in biogeography. *Applied Statistics*, 43:569–582.

Heikkinen, J. and Penttinen, A. (1999). Bayesian smoothing in the estimation of the pair potential function of Gibbs point processes. *Bernoulli*, in press.

Hoeting, J. A., Van Caster, M., and Bowden, D. (2000). An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics*, 5:102–114.

Högmander, H. and Møller, J. (1995). Estimating distribution maps from atlas data using methods of statistical image analysis. *Biometrics*, 51:393–404.

Huffer, F. W. and Wu, H. L. (1998). Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics*, 54:509–524.

Hughes, J. P. and Guttorp, P. (1994). Incorporating spatial dependence and atmospheric data in a model of precipitation. *Journal of Applied Meteorology*, 33:1503–1515.

Ickstadt, K. and Wolpert, R. L. (1996). Spatial correlation or spatial variation? A comparison of Gamma/Poisson hierarchical models. Technical report, Institute of Statistics and Decision Sciences, Duke University, Durham NC 27708-0251.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31:253–258.

Johnson, N. L., Kotz, S., and Kemp, A. W. (1992). *Univariate Discrete Distributions*. John Wiley & Sons, New York.

Kaiser, M. S. and Cressie, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statistics and Probability Letters*, 35:423–432.

Lee, J. and Kaiser, M. S. (1997). Maximum likelihood estimation in conditionally specified statistical models. Technical report, Department of Statistics, Iowa State University.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society, B*, 61:381–400.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrica*, 37:17–23.

Penttinen, A. (1984). Modelling interaction in spatial point patterns: Parameter estimation by the maximum likelihood method. *Jyvaskyla Studies in Computer Science, Economics, and Statistics*, pages 1–107.

Preisler, H. K. (1993). Modelling spatial patterns of trees attacked by bark-beetles. *Applied Statistics*, 42:501–514.

Preisler, H. K. and Mitchell, R. G. (1993). Colonization patterns of the mountain pine-beetle in thinned and unthinned lodgepole pine stands. *Forest Science*, 39:528–545.

Wang, Y. (1998). Smoothing spline models with correlated errors. *Journal of the American Statistical Association*, 93:341–348.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41:434–439.

Wolpert, R. L. and Ickstadt, K. (1998). Gamma/Poisson random field models for spatial statistics. *Biometrika*, 85:251–267.

Wood, S. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society, Series B*, 62(1):413–428.

Wu, H. L. and Huffer, F. W. (1997). Modeling the distribution of plant species using the autologistic regression model. *Environmental and Ecological Statistics*, 4:49–64.

Younes, L. (1988). Estimation and annealing for Gibbsian fields. *Annales de l'Institut Henri Poincaré*, 24:269–294.

Table 1: Means of parameter estimates, using pseudo likelihood (MPL), coding (CODE1, CODE2) and Monte Carlo maximum likelihood (MCML)), from 100 simulated data sets for each of simulation *studies 1-3.* The mean of asymptotic standard errors are given in brackets, empirical standard errors are given in square brackets.

Parameter	α	(s. e.)	[s.e.]	bias	γ	(s. e.)	[s.e.]	bias
Study 1								
($r = 100$)								
True values	*1.609*				*0*			
MPL	1.594	(0.131)	[0.187]	-0.016	0.003	(0.026)	[0.036]	0.003
CODE1	1.578	(0.189)	[0.201]	-0.031	0.006	(0.037)	[0.038]	0.006
CODE2	1.568	(0.188)	[0.189]	-0.042	0.008	(0.037)	[0.037]	0.008
MCML	1.592	(0.175)	[0.187]	-0.017	0.003	(0.034)	[0.036]	0.003
Study 2								
($r = 20$)								
True values	*1.609*				*0.050*			
MPL	1.640	(0.128)	[0.173]	0.030	0.045	(0.017)	[0.024]	-0.005
CODE1	1.618	(0.183)	[0.178]	0.009	0.048	(0.025)	[0.024]	-0.002
CODE2	1.611	(0.185)	[0.194]	0.002	0.049	(0.025)	[0.027]	-0.001
MCML	1.644	(0.157)	[0.172]	0.035	0.045	(0.021)	[0.024]	-0.005
Study 3								
($r = 100$)								
True values	*1.609*				*-0.500*			
MPL	1.600	(0.091)	[0.082]	-0.010	-0.499	(0.051)	[0.050]	0.001
CODE1	1.590	(0.135)	[0.136]	-0.019	-0.497	(0.074)	[0.077]	0.003
CODE2	1.606	(0.131)	[0.128]	-0.004	-0.503	(0.074)	[0.075]	-0.003
MCML	1.585	(0.074)	[0.076]	-0.024	-0.491	(0.042)	[0.042]	0.009

Table 2: Means of parameter estimates from *study 4*, using pseudo likelihood (MPL), coding (CODE1, CODE2) and Monte Carlo maximum likelihood (MCML)), from 100 simulated data sets. The mean of asymptotic standard errors are given in brackets, empirical standard errors are given in square brackets.

Parameter	α	(s. e.)	[s.e.]	bias	β	(s. e.)	[s.e.]	bias	γ	(s. e.)	[s.e.]	bias
Study 4												
($r = 100$)												
True values	*1.609*				*0.050*				*0*			
MPL	1.603	(0.130)	[0.178]	-0.006	0.050	(0.009)	[0.011]	0.000	0.002	(0.025)	[0.033]	0.002
CODE1	1.567	(0.188)	[0.190]	-0.042	0.048	(0.013)	[0.013]	-0.002	0.008	(0.036)	[0.034]	0.008
CODE2	1.567	(0.190)	[0.185]	-0.043	0.048	(0.014)	[0.014]	-0.002	0.009	(0.036)	[0.036]	0.009
MCML	1.601	(0.167)	[0.173]	-0.008	0.050	(0.011)	[0.011]	0.000	0.002	(0.032)	[0.032]	0.002

Table 3: Mite example parameter estimates. Results for MCNR are taken from Lee and Kaiser (1997). X^2 is the Pearson statistic of goodness-of-fit.

parameter	MCML	MCNR	MPL	ML
X^2_{df}	72.18_{62}	-	71.96_{62}	74.62_{61}
$\hat{\alpha}$	-0.199	-0.206	-0.215	0.198
s.e.	0.270	0.270	0.239	0.113
$\hat{\gamma}$	0.087	0.088	0.090	–
s.e.	0.051	0.052	0.043	–

Table 4: Results for the four different models (Poisson, auto-Poisson, NegBin and auto-NegBin) fitted to seed counts of *Poa* spp (year 1): Residual deviance, *df*, estimates of the coefficient of autocorrelation $\hat{\gamma_{ij}}$ and aggregation parameter $\hat{k}$ (fixed for the NegBin model). All models have as a minimum terms for block, weed, crop, cult and the weed.cult interaction. For MCML deviance residuals and Pearson's residuals for X^2 and Moran's correlation coefficient are defined as in section 4.

	Poisson	auto Poisson MPL	auto Poisson MCML	Negative Binomial	auto NegBin MPL
df	66	65	65	66	65
X^2	100.2	-	95.2	69.1	-
residual deviance	111.7	102.9	112.4	83.8	79.5
$\hat{ns}$ (north/south)	n.a.	0.169	0.034	n.a.	0.165
{asymptotic s.e.}		{0.072}	{0.056}		{0.075}
$\hat{ew}$ (east/west)	n.a.	-	-	n.a.	-
$\hat{k}$ (aggregation)	n.a.	n.a.	n.a.	3.3	3.3
{asymptotic s.e.}	-	-	-	2.07	n.a.
Can residuals be assumed random in space?	yes	-	yes	yes	-
Moran corr. (first order)	0.047	-	0.046	0.047	-
(p-value perm. test.)	(0.247)	-	(0.209)	(0.238)	-
Can residuals be assumed random in space?	yes	-	yes	yes	-
Moran corr. (north/south)	0.151	-	0.147	0.121	-
(p-value perm. test.)	(0.085)	-	(0.100)	(0.138)	-

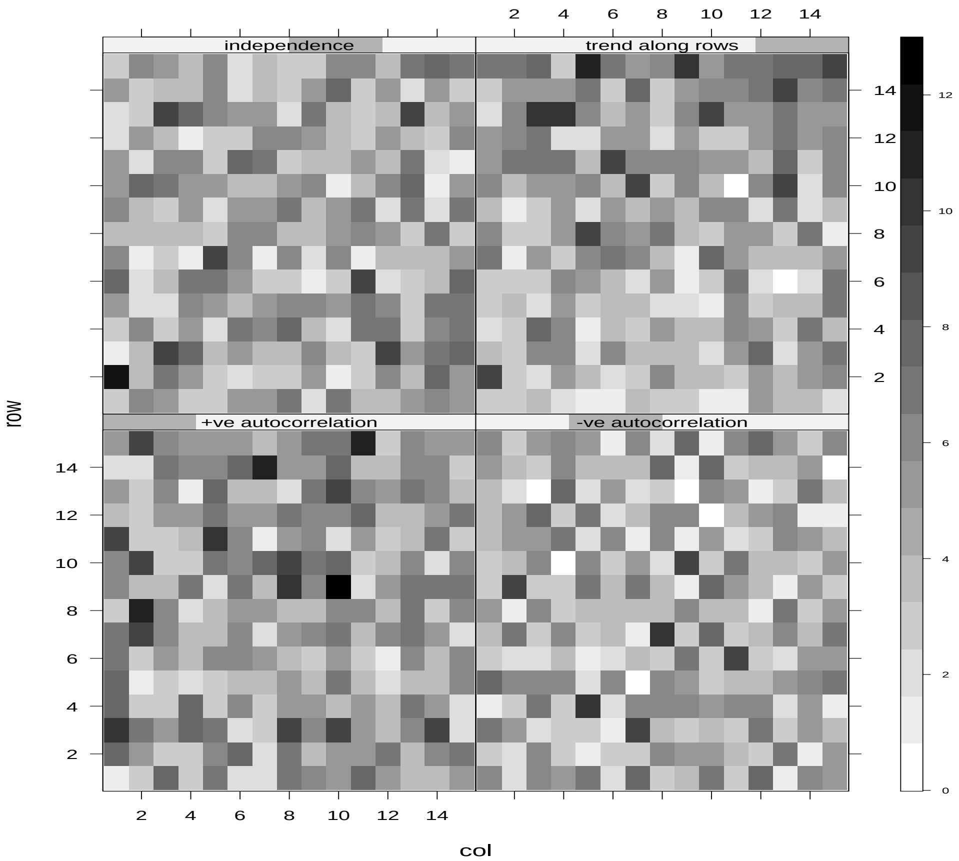


Figure 1: Simulated counts from models assuming different degrees of autocorrelation.

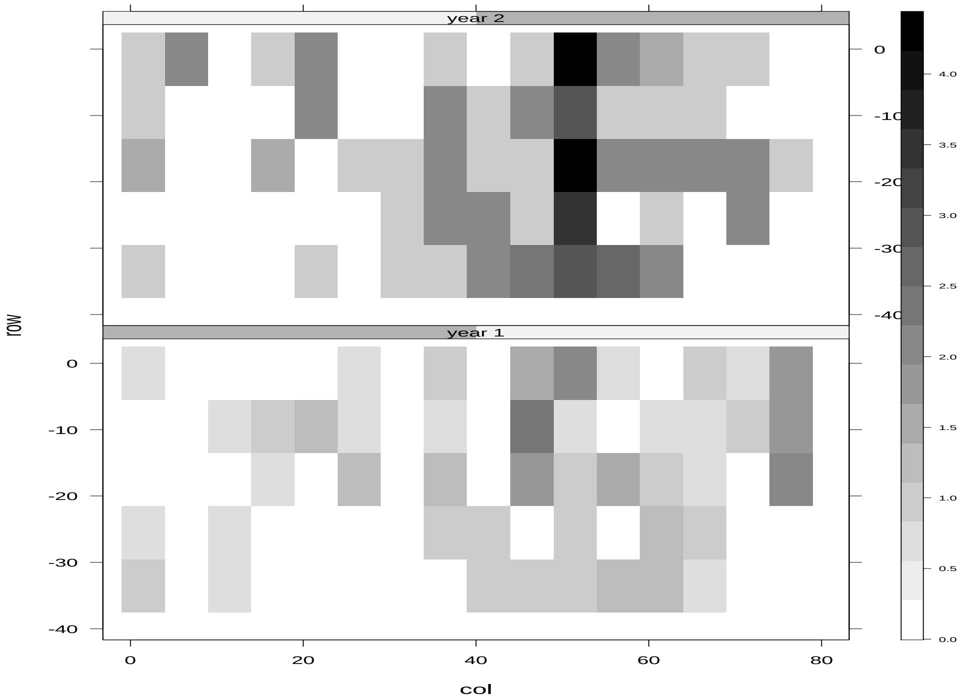


Figure 2: Observed seed counts (log(count +1)) of the weed *Poa* sp. in an agricultural experiment observed in two consecutive years.

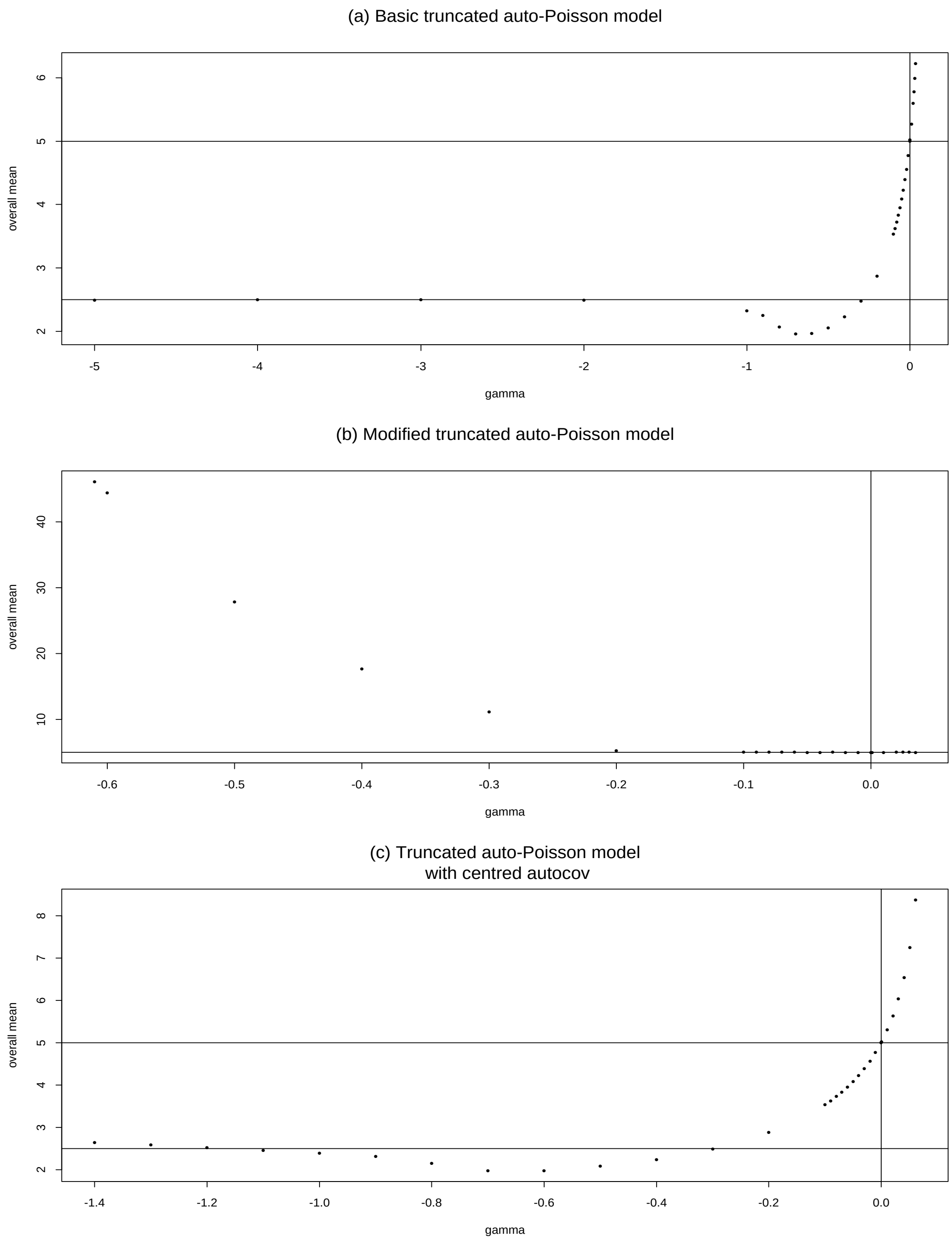


Figure 3: Average of $\bar{y}$ from 500 realisations from (a) the *basic truncated auto-Poisson model* with parameters $\alpha = \log(5)$ and different values of γ; (b) the *modified truncated auto-Poisson model* with parameters as in (a), (c) the *truncated auto-Poisson model* with a centred *autocovariate*, parameters as in (a).

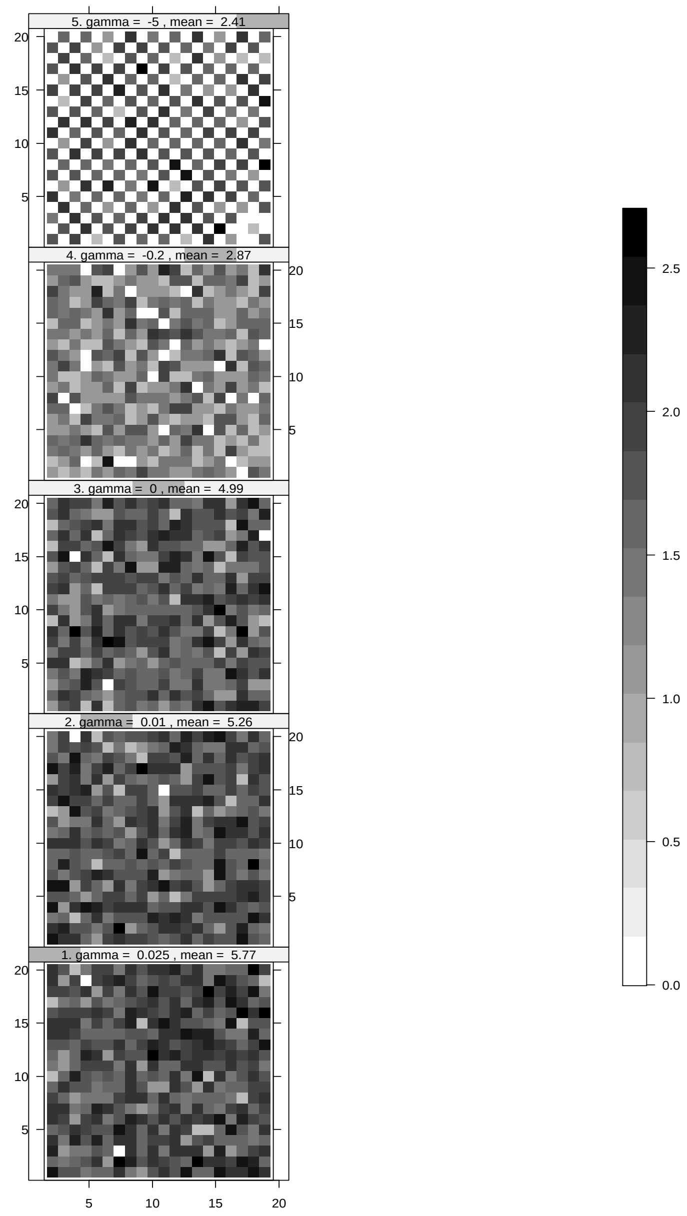


Figure 4: log(count+1) from single realisations using the *basic truncated auto-Poisson model* with $\alpha = \log(5)$ and $\gamma = -5, -0.5, 0, 0.01, 0.02$ (from top to bottom) applied to a first-order neighbourhood.

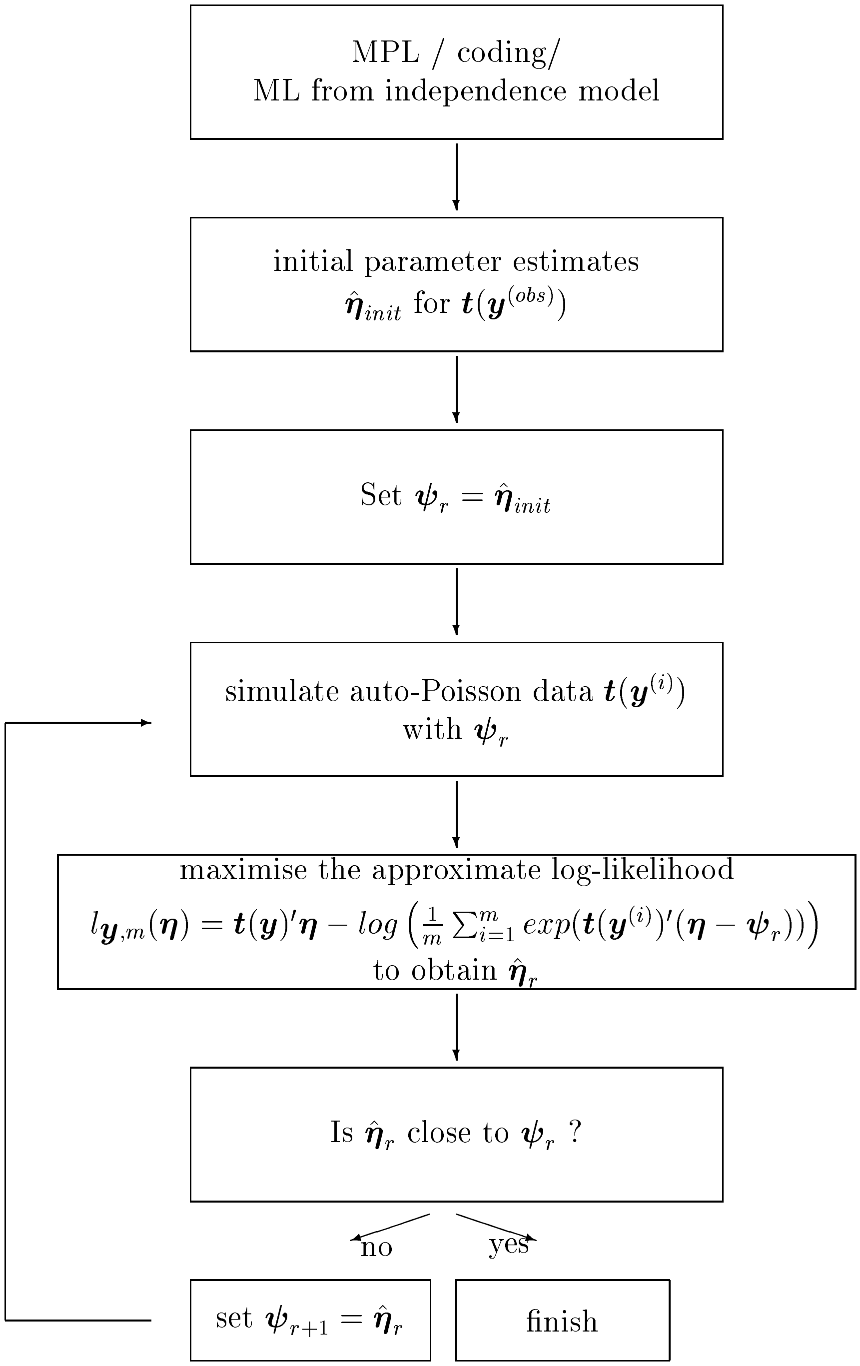


Figure 5: Flowchart for Monte Carlo maximum likelihood (MCML). The vector $\boldsymbol{t}(\boldsymbol{y}^{(obs)})$ refers to summary statistics from the observed data and $\boldsymbol{t}(\boldsymbol{y}^{(i)})$ to summary statistics for simulated data.

Figure 6: Observed mite counts.

```
2 1 2 1 0 0 1 2
1 1 1 1 3 4 1 4
0 1 0 2 2 1 3 1
0 0 0 3 3 0 1 2
2 1 0 1 1 1 0 0
1 1 0 1 2 1 0 1
0 3 1 0 1 3 3 3
0 0 0 0 1 5 0 1.
```

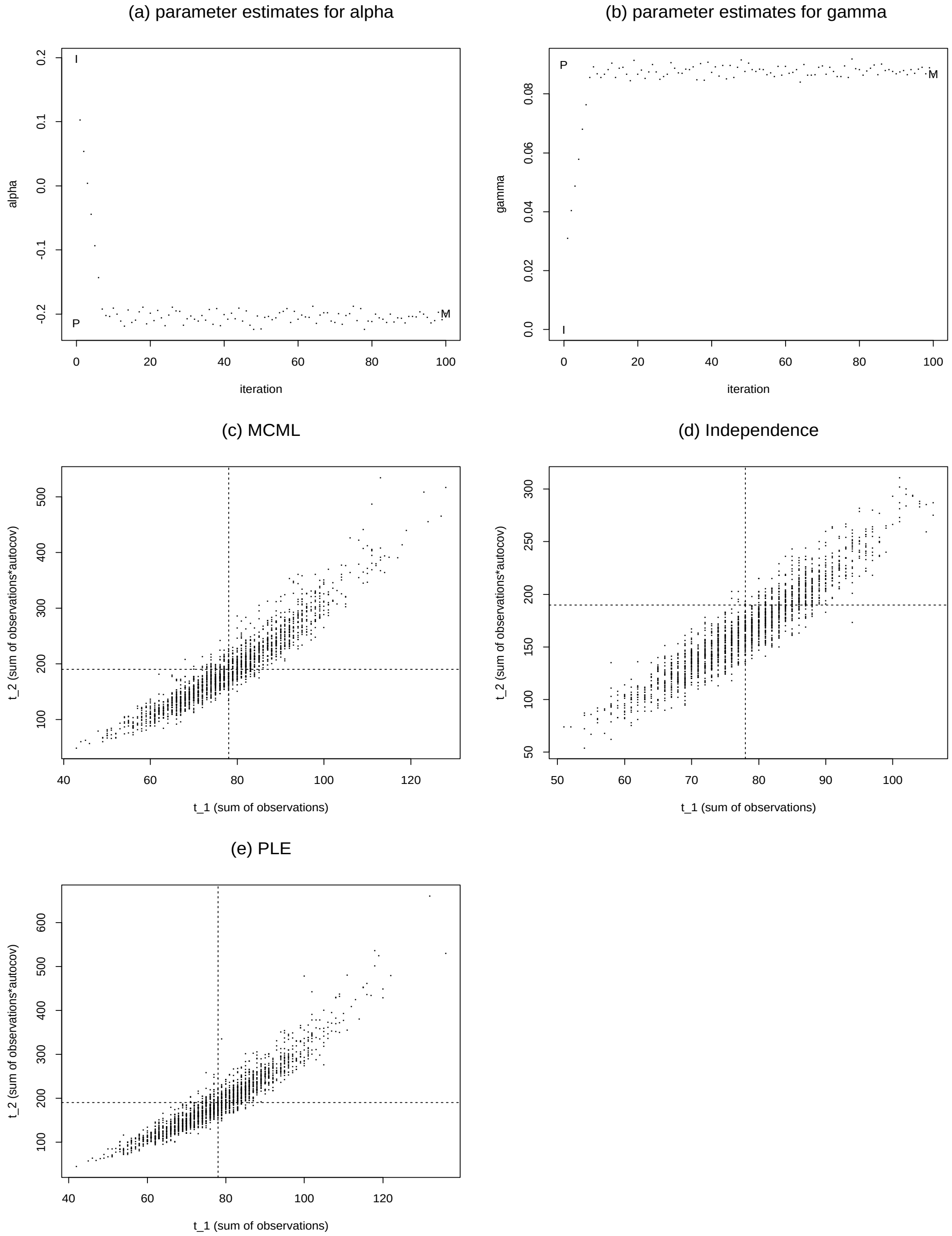


Figure 7: Mite example: parameter estimates for (a) α and (b) γ for the iterative process of MCML (marked as dots, final estimate as M) starting from ML estimates of the independence model (marked as I) and parameter estimates for MPL (marked as P). (c) - (e) are plots of simulated summary statistics $t_1^{(t)}$ versus $t_2^{(t)}$ using parameters from MCML, the independence model and MPL. Where the two dotted lines cross corresponds to observed t_1 and t_2.

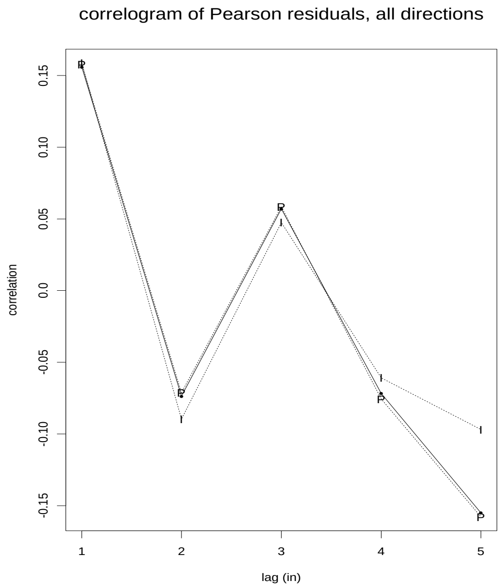


Figure 8: Omni-directional correlograms of Pearson's residuals from the auto-Poisson (MCML - solid line, MPL - dotted line) and independence (ML - dashed line) model fitted to the mite counts.

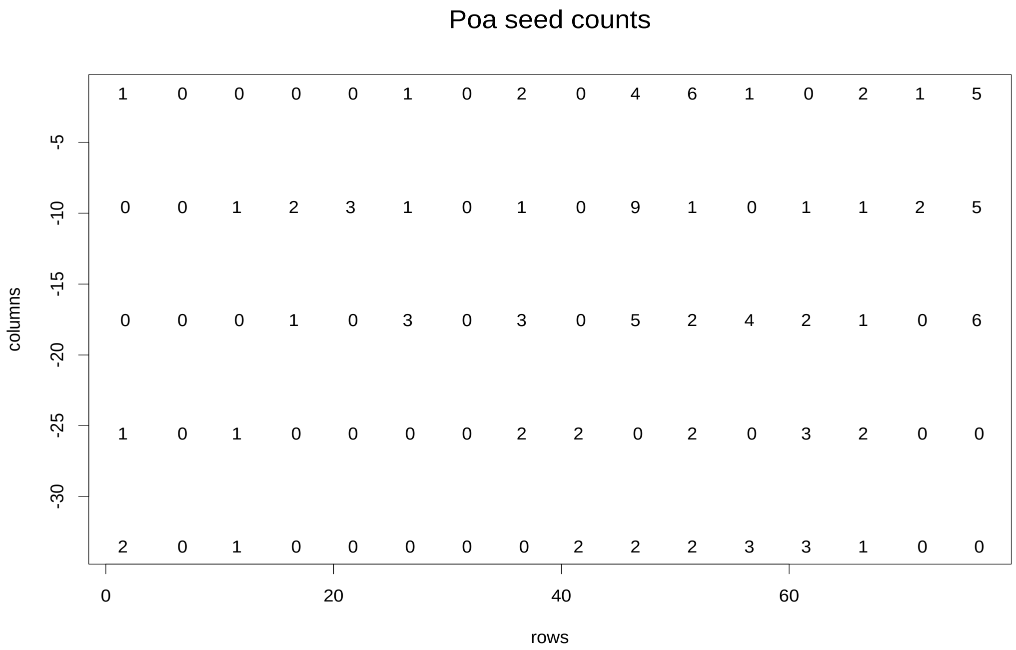
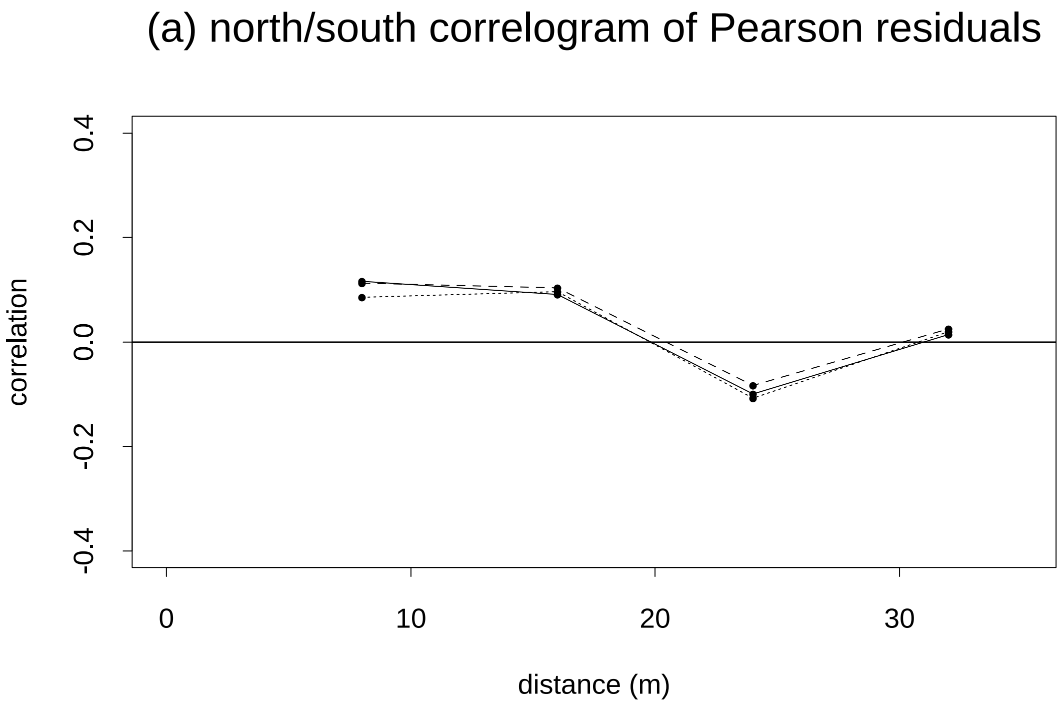


Figure 9: *Poa* seed counts sample in soil cores of volume 1.2 l.



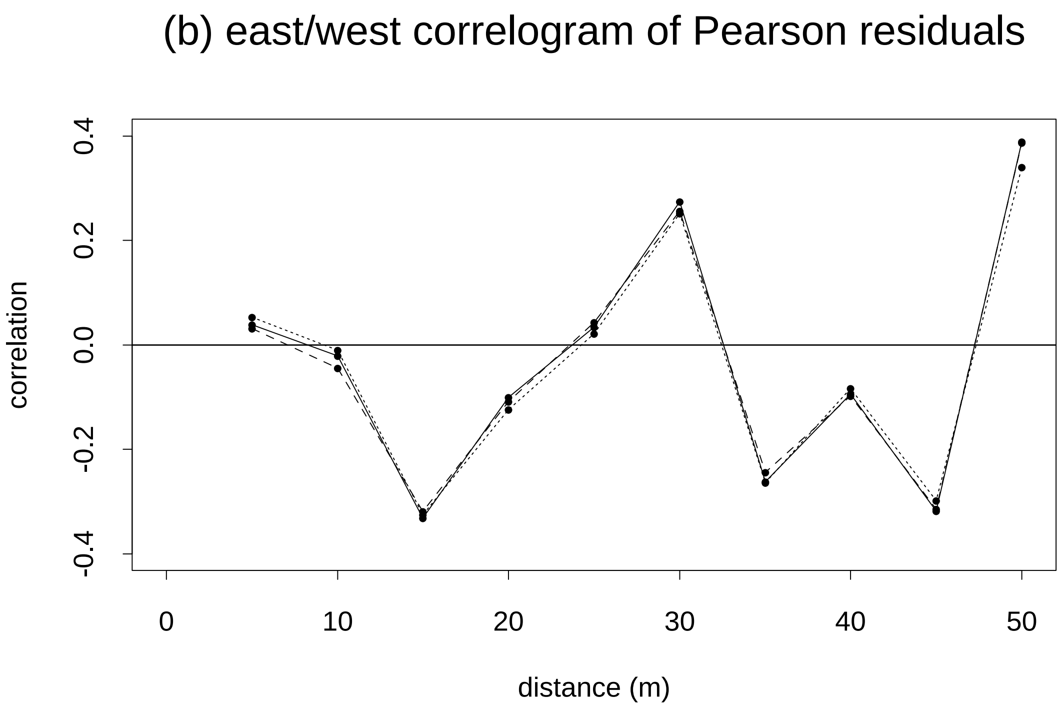


Figure 10: Correlograms in (a) north/south direction and (b) east/west direction of Pearson residuals from Poisson (solid line), auto-Poisson (dotted line) and NegBin model (dashed line) fitted to *Poa* year 1 seed counts. The auto-Poisson model is fitted by MCML.