

# Modelling count data with overdispersion and spatial effects

Susanne Gschlößl<sup>1</sup>, Claudia Czado<sup>2</sup>

<sup>1</sup> Both at Center of Mathematical Sciences, Munich University of Technology, Boltzmannstr.3 , D-85747 Garching, Germany, e-mail: [susanne@ma.tum.de](mailto:susanne@ma.tum.de)

<sup>2</sup> e-mail: [cczado@ma.tum.de](mailto:cczado@ma.tum.de)

Received: date / Revised version: date

**Abstract** In this paper we consider regression models for count data allowing for overdispersion in a Bayesian framework. We account for unobserved heterogeneity in the data in two ways. On the one hand, we consider more flexible models than a common Poisson model allowing for overdispersion in different ways. In particular, the negative binomial and the generalized Poisson distribution are addressed where overdispersion is modelled by an additional model parameter. Further, zero-inflated models in which overdispersion is assumed to be caused by an excessive number of zeros are discussed. On the other hand, extra spatial variability in the data is taken into account by adding spatial random effects to the models. This approach allows for an underlying spatial dependency structure which is modelled using a conditional autoregressive prior based on Pettitt et al. (2002). In an application the presented models are used to analyse the number of invasive meningococcal disease cases in Germany in the year 2004. Models are compared according to the deviance information criterion (DIC) suggested by Spiegelhalter et al. (2002) and using proper scoring rules, see for example Gneiting and Raftery (2004). We observe a rather high degree of overdispersion in the data which is captured best by the GP model when spatial effects are neglected. While the addition of spatial effects to the models allowing for overdispersion gives no or only little improvement, a spatial Poisson model is to be preferred over all other models according to the considered criteria.

## 1 Introduction

A very popular model for count data is the Poisson distribution. However, in the Poisson model equality of the variance and the mean is assumed which is too restrictive for overdispersed data where the variance in the data is higher than the expected one from the model. This paper aims to give an overview over different models for spatially indexed count data allowing for overdispersion in a Bayesian perspective. We follow two approaches for dealing with the extra variability in overdispersed data. On the one hand, we consider a wider class of models allowing for overdispersion, on the other hand spatial random effects are introduced to capture unobserved spatial heterogeneity in the data.

Overdispersion with respect to the Poisson model can be modelled by introducing an additional parameter. In particular we consider the negative binomial (NB) distribution and the generalized Poisson (GP) distribution introduced by Consul and Jain (1973). Both models allow an independent modelling of the mean and the variance by the inclusion of an additional parameter.

When dealing with a data set with an excessive number of zeros, zero-inflated models might be used, see for example Winkelmann (2003). In contrast to the GP and the NB model, overdispersion in zero inflated models is caused by the occurrence of more zero observations than expected. Zero inflated models can be used in combination with any model for count data. Additionally to the zero observations arising from the count data model an extra proportion of zeros is incorporated. Lambert (1992) introduced the zero inflated Poisson regression model, a Bayesian analysis of the zero inflated Poisson model is given in Rodrigues (2003). Zero inflated regression models in combination with the generalized Poisson distribution have been addressed in Famoye and Singh (2003b) and Famoye and Singh (2003a) using maximum likelihood estimation, a Bayesian analysis without the inclusion of covariates is given in Angers and Biswas (2003). Agarwal et al. (2002) use a zero inflated Poisson regression model for spatial count data in a Bayesian framework.

The second approach for modelling unobserved data heterogeneity is the introduction of random effects. For spatially indexed data which are the focus in this paper, spatial random effects associated with each region or site may be used, allowing for the modelling of an underlying spatial dependency structure.

In this paper, we consider Poisson, NB, GP and zero inflated (ZI) regression models both including and without spatial random effects in a Bayesian context. In contrast to classical inference the Bayesian approach allows to adjust for parameter uncertainty by assigning prior distributions to the parameters. Further, a spatial correlation structure is easily incorporated in a Bayesian setting by assuming an adequate prior distribution for the spa-

tial random effects. In this paper we assume a proper Gaussian conditional autoregressive spatial prior based on Pettitt et al. (2002) which takes the neighbourhood structure of the data into account and allows for spatial dependencies. Since this results in a high dimensional, complex posterior distribution, Markov Chain Monte Carlo (MCMC) is used for parameter estimation.

We give an application of the considered models to disease mapping. In particular, we analyse the number of invasive meningococcal disease cases registered in Germany in the year 2004. Models are compared using the deviance information criterion (DIC) suggested by Spiegelhalter et al. (2002) and proper scoring rules, see for example Gneiting and Raftery (2004). We observe a substantial degree of overdispersion in the data which is modelled best by the GP distribution when spatial effects are neglected. While the addition of spatial random effects gives no or little improvement to the models allowing for overdispersion, spatial effects turn out to be significant for the Poisson model. In particular, according to the DIC and the scoring rules a spatial Poisson models gives the best fit for these data. However, no smooth spatial pattern is modelled. Instead some isolated regions with high risk are detected by the spatial effects, indicating that the risk is not sufficiently explained by the incorporated covariates in these regions.

This paper is organized as follows. In Section 2 the negative binomial, the generalized Poisson and zero-inflated regression models are presented. The conditional autoregressive prior assumed for the spatial effects is discussed in Section 3, prior assumptions for the regression and model dependent overdispersion parameters are given in Section 4. The DIC and the used proper scoring rules are reviewed in Section 5. Finally, in Section 6 the presented models are applied to analyse the number of invasive meningococcal disease cases in Germany is given. Section 7 gives a summary of the results and draws conclusions. Details about the MCMC algorithms can be found in the Appendix.

## 2 Models for count data including overdispersion

A commonly used model for count data is the Poisson model, where equality of mean and variance is assumed. Since this condition is not satisfied any more if overdispersion is present in the data, we consider models, which allow the variance to be larger than the mean in this section. For a detailed study of various count data models see Winkelmann (2003).

### 2.1 Negative Binomial (NB) distribution and Regression

The density of the negative binomial distribution with parameters  $r > 0$  and  $\mu > 0$  denoted by  $\text{NB}(r, \mu)$  is defined by

$$P(Y = y|r, \mu) = \frac{\Gamma(y+r)}{\Gamma(r)y!} \cdot \left(\frac{r}{\mu+r}\right)^r \cdot \left(\frac{\mu}{\mu+r}\right)^y, \quad y = 0, 1, 2, \dots \quad (2.1)$$

with

$$E(Y|r, \mu) = \mu \quad \text{and} \quad \text{Var}(Y|r, \mu) = \mu \left(1 + \frac{\mu}{r}\right).$$

The variance is the mean multiplied by the positive factor  $\varphi := 1 + \frac{\mu}{r}$  and therefore greater than the mean, i.e. overdispersion can be modelled in the negative binomial distribution. We call the factor  $\varphi$  dispersion factor. In the limit  $r \rightarrow \infty$  the NB distribution converges to the Poisson distribution with parameter  $\mu$ , see Winkelmann (2003). The negative binomial distribution also arises from a Poisson distribution where the parameter  $\theta$  is assumed to be random and to follow a Gamma distribution with mean  $E(\theta) = \mu$  and variance  $\text{Var}(\theta) = \frac{\mu^2}{r}$ . Therefore, overdispersion in the NB model can be interpreted by unobserved heterogeneity among observations. In a regression model with  $Y_i \sim \text{NB}(r, \mu_i)$  independent for  $i = 1, \dots, n$ , the mean of  $Y_i$  is specified in terms of covariates  $\mathbf{x}_i$  and unknown regression parameters  $\boldsymbol{\beta}$  by

$$E(Y_i|\mathbf{x}_i, \boldsymbol{\beta}) = \mu_i > 0.$$

Note, that in the NB regression model the dispersion factor  $\varphi_i := 1 + \frac{\mu_i}{r}$  takes observation specific values.

### 2.2 Generalized Poisson (GP) distribution and Regression

The generalized Poisson distribution has been introduced by Consul and Jain (1973) and is investigated in detail in Consul (1989). A random variable  $Y$  is called generalized Poisson distributed with parameters  $\mu > 0$  and  $\lambda$ , denoted by  $\text{GP}(\mu, \lambda)$ , if

$$P(Y = y|\mu, \lambda) = \begin{cases} \mu[\mu(1-\lambda) + \lambda y]^{y-1} \frac{(1-\lambda)}{y!} \exp[-\mu(1-\lambda) - \lambda y], & y = 0, 1, 2, \dots \\ 0 & \text{for } y > m \text{ when } \lambda < 0 \end{cases} \quad (2.2)$$

where  $\max(-1, -\frac{\mu}{m-\mu}) < \lambda < 1$  and  $m(\geq 4)$  is the largest positive integer for which  $\mu(1-\lambda) + m\lambda > 0$  for negative  $\lambda$ . Mean and variance are given by

$$E(Y|\mu, \lambda) = \mu \quad \text{and} \quad \text{Var}(Y|\mu, \lambda) = \frac{\mu}{(1-\lambda)^2}, \quad (2.3)$$

hence  $\varphi := \frac{1}{(1-\lambda)^2}$  can be interpreted as an dispersion factor for the GP distribution. For  $\lambda = 0$ , the generalized Poisson distribution reduces to the

Poisson distribution with parameter  $\mu$ , for  $\lambda < 0$  underdispersion can be modelled, whereas for  $\lambda > 0$  overdispersion is obtained. The focus in this paper is the modelling of overdispersion, therefore  $\lambda$  is assumed to take only values in the interval  $[0, 1)$  in the remainder of this paper. Similar to the NB model, the GP distribution is a mixture of Poisson distributions as has been proved by Joe and Zhu (2005). A regression model for independent GP( $\mu_i, \lambda$ ) distributed response variables  $Y_i, i = 1, \dots, n$  is set up by specifying the mean by

$$E(Y_i | \mathbf{x}_i; \boldsymbol{\beta}, \lambda) = \mu_i > 0$$

like in the NB model. While the dispersion parameter in the NB regression model depends on  $\mu_i$  leading to a variance function which is quadratic in  $\mu_i$ , the dispersion parameter  $\varphi = \frac{1}{(1-\lambda)^2}$  in the GP regression model is the same for each observation and results in a linear variance function.

### 2.3 Comparison of NB and GP distribution

In order to compare the behaviour of the NB and the GP distribution, we equate the mean and the variance of a GP( $\mu, \lambda$ ) with the mean and the variance of a NB( $r, \mu$ ) distributed random variable, i.e.

$$\frac{\mu}{(1-\lambda)^2} = \mu \left(1 + \frac{\mu}{r}\right)$$

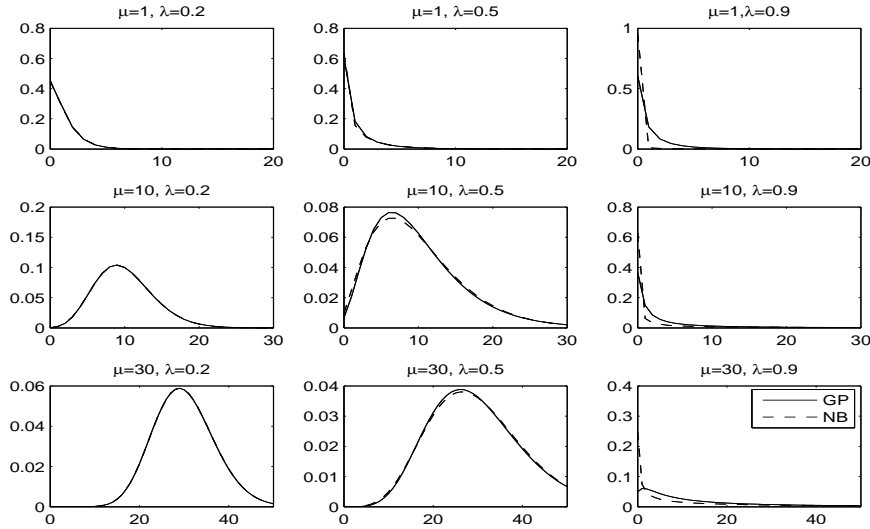
has to hold and the equation

$$r = \frac{\mu(1-\lambda)^2}{\lambda(2-\lambda)} \quad (2.4)$$

is obtained. In Figure 1 the NB distribution is plotted in comparison to the GP distribution with  $\mu$  and  $r$  chosen according to (2.4). For a better visual comparison the densities of these discrete distributions are presented as line plots. For small values of  $\lambda$  both distributions behave very similarly. With increasing values of  $\lambda$  slight differences between the two distributions can be observed which become greater when  $\lambda$  tends to 1. In particular, the negative binomial distribution gives more mass to small values of  $y$  if strong overdispersion is present.

### 2.4 Zero Inflated (ZI) Models

For count data with an excessive number of zero observations zero inflated (ZI) models can be used. These models allow for a higher number of zeros than can be explained by standard models for count data. Additional to the zero observations arising from the supposed count data distribution, a proportion of extra zeros is assumed. ZI models have been widely used in the literature, for a short overview see Winkelmann (2003).



**Fig. 1** Comparison of the generalized Poisson distribution with  $\mu = 1, 10, 30$  and  $\lambda = 0.2, 0.5, 0.9$  to the negative binomial distribution with  $\mu = 1, 10, 30$  and  $r = \frac{\mu(1-\lambda)^2}{\lambda(2-\lambda)}$ .

Let  $\pi(y|\boldsymbol{\theta})$  be a distribution function for count data with unknown parameters  $\boldsymbol{\theta}$ . Then a zero inflated model with extra proportion  $p \in [0, 1]$  of zeros is defined by (see Agarwal et al. (2002))

$$P(Y = y|p, \boldsymbol{\theta}) = \begin{cases} p + (1-p)\pi(y=0|\boldsymbol{\theta}) & \text{if } y = 0 \\ (1-p)\pi(y|\boldsymbol{\theta}) & \text{if } y > 0 \end{cases} \quad (2.5)$$

Mean and variance are given by

$$E(Y|p, \boldsymbol{\theta}) = (1-p)E_{\pi}(Y|\boldsymbol{\theta}) \quad (2.6)$$

and

$$Var(Y|p, \boldsymbol{\theta}) = p(1-p)[E_{\pi}(Y|\boldsymbol{\theta})]^2 + (1-p)Var_{\pi}(Y|\boldsymbol{\theta}). \quad (2.7)$$

The introduction of latent indicator variables  $\mathbf{Z} = (Z_1, \dots, Z_n)'$  leads to a model which is easier to handle in a Bayesian context and in particular allows a Gibbs step for  $p$ .  $Z_i$  takes the value  $z_i = 0$  for all observations with  $y_i > 0$ . For all zero observations  $y_i = 0$ , the latent variable takes the value  $z_i = 0$  if observation  $i$  arises from the count data distribution  $\pi(y|\boldsymbol{\theta})$  and the value  $z_i = 1$  for extra zeros. Marginally,  $Z_i \sim \text{Bernoulli}(p_i)$ . Using the latent variables  $\mathbf{Z}$ , the joint likelihood of  $\mathbf{Y} = (Y_1, \dots, Y_n)'$  and  $\mathbf{Z}$  is given by

$$\begin{aligned} f(\mathbf{Y}, \mathbf{Z}|p, \boldsymbol{\theta}) &= \prod_{i=1}^n p_i^{z_i} [(1-p_i)\pi(y_i|\boldsymbol{\theta})]^{1-z_i} \\ &= \prod_{i:y_i=0} p_i^{z_i} [(1-p_i)\pi(0|\boldsymbol{\theta})]^{1-z_i} \cdot \prod_{i:y_i>0} (1-p_i)\pi(y_i|\boldsymbol{\theta}). \end{aligned}$$

In this paper we will focus on the zero inflated Poisson and the zero inflated generalized Poisson models, which are special cases of the ZI model (2.5). The zero inflated negative binomial distribution will not be discussed in this paper, since the GP model turned out to be more adequate than the NB model for the application considered later on.

#### 2.4.1 Zero Inflated Poisson (ZIP) Distribution

Here the Poisson distribution is assumed for the underlying count data distribution, i.e.  $\pi(y|\theta) := \pi(y|\mu) = \mu^y \frac{\exp(-\mu)}{y!}$ . Mean and variance of the ZIP distribution, denoted by  $ZIP(p, \mu)$ , are specified by  $E(Y|p, \mu) = (1 - p)\mu$  and  $Var(Y|p, \mu) = (1 - p)\mu(\mu p + 1) = E(Y|p, \mu)(\mu p + 1)$ . For  $p > 0$  the dispersion factor  $\varphi := \mu p + 1$  of the ZIP model is positive, i.e. the presence of extra zeros leads to overdispersion.

#### 2.4.2 Zero Inflated Generalized Poisson (ZIGP) Distribution

The ZIGP regression model was already introduced by Famoye and Singh (2003b), in Famoye and Singh (2003a) a generalisation to k-inflated GP regression models is given. The ZIGP distribution, denoted by  $ZIGP(p, \mu, \lambda)$ , is obtained if the density function of the GP distribution given in (2.2) is chosen for  $\pi(y|\theta)$ . The mean and the variance of the ZIGP distribution are given by  $E(Y|p, \mu, \lambda) = (1 - p)\mu$  and  $Var(Y|p, \mu, \lambda) = E(Y|p, \mu, \lambda) \left[ p\mu + \frac{1}{(1-\lambda)^2} \right]$ . The dispersion factor of the ZIGP model is therefore given by  $\varphi := p\mu + \frac{1}{(1-\lambda)^2}$ . Here, overdispersion can both result from the overdispersion parameter  $\lambda$  of the GP distribution and the extra proportion of zeros  $p$  when  $p > 0$ .

#### 2.4.3 Zero Inflated Regression Models

In a regression model  $Y_i \sim ZIP(p_i, \mu_i)$  and  $Y_i \sim ZIGP(p_i, \mu_i, \lambda)$ , independent for  $i = 1, \dots, n$ , respectively, a regression can be performed both for  $\mathbf{p} = (p_1, \dots, p_n)'$  and  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ . As in the NB and GP model the parameter  $\mu_i$  is assumed to depend on covariates  $\mathbf{x}_i$  and unknown regression parameters  $\boldsymbol{\beta}$ . For the proportion of extra zeros a logistic link might be chosen, i.e.  $p_i = \frac{\exp(\tilde{\mathbf{x}}_i' \boldsymbol{\alpha})}{1 + \exp(\tilde{\mathbf{x}}_i' \boldsymbol{\alpha})}$  with covariate vector  $\tilde{\mathbf{x}}_i$  and regression parameters  $\boldsymbol{\alpha}$ . However, for the data considered in our application no significant zero inflation is detected. An extension to ZI models with regression on  $p$  therefore seems unnecessary and is not addressed further in this paper.

Alternatively to ZI models hurdle models, see for example Winkelmann (2003) for an overview, could be used. The most widely used hurdle model is the zero hurdle model which assumes two separate models for zero and non-zero observations. Attention is however restricted to zero inflated models in this paper. As mentioned above, the amount of zeros in the data analysed in Section 6 turned out to be covered sufficiently well by non-zero inflated models and gives no rise for a separate analysis of zero observations.

### 3 Spatial effects using a Gaussian conditional autoregressive model

In addition to covariates we will incorporate spatial random effects in the regression models in order to account for spatial heterogeneity as well as spatial correlation in the data. We consider models for data aggregated in regions. A spatial dependency structure is imposed by assuming a prior distribution for the spatial effects which takes the neighbourhood structure of the area under consideration into account. In particular we consider a special case of the Gaussian conditional autoregressive (CAR) model introduced by Pettitt et al. (2002). Assume the data to be distributed on  $J$  regions  $\{1, \dots, J\}$ . Then the vector  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_J)'$  of spatial effects is assumed to follow a multivariate normal distribution, in particular

$$\boldsymbol{\gamma} \sim N(0, \sigma^2 Q^{-1}) \quad (3.1)$$

where the elements of the precision matrix  $Q = (Q_{ij})$ ,  $i, j = 1, \dots, J$  are given by

$$Q_{ij} = \begin{cases} 1 + |\psi| \cdot N_i & i = j \\ -\psi & i \neq j, i \sim j \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

We write  $i \sim j$  for regions  $i$  and  $j$  which are contiguous and assume regions to be neighbours if they share a common border.  $N_i$  denotes the number of neighbours of region  $i$ . The conditional distribution of  $\gamma_i$ , given all the remaining components  $\boldsymbol{\gamma}_{-i}$ ,  $i = 1, \dots, J$  is given by

$$\gamma_i | \boldsymbol{\gamma}_{-i} \sim N\left(\frac{\psi}{1 + |\psi| \cdot N_i} \sum_{j \sim i} \gamma_j, \sigma^2 \frac{1}{1 + |\psi| \cdot N_i}\right).$$

The parameter  $\psi$  determines the overall degree of spatial dependence, for  $\psi = 0$  all regions are spatially independent, whereas for  $\psi \rightarrow \infty$  the degree of dependence increases. Pettitt et al. (2002) show that  $Q$  is symmetric and positive definite, therefore (3.1) is a proper distribution. Another convenient feature of this CAR model is that the determinant of  $Q$  which is needed for the update of  $\psi$  in a Markov Chain Monte Carlo (MCMC) algorithm can be computed efficiently, see Pettitt et al. (2002) for details. Many other authors have dealt with conditional autoregressive models. An overview about CAR models is given in the book by Banerjee et al. (2004) and in Jin et al. (2004) where also multivariate CAR models are discussed. The most popular model is probably the intrinsic CAR model introduced by Besag and Kooperberg (1995). The joint density for  $\boldsymbol{\gamma}$  in the intrinsic CAR model is improper in contrast to model (3.3) described above which has a proper joint density. Czado and Prokopenko (2004) consider a modification of model (3.3) which is a proper model as well but reduces to the intrinsic CAR model in the limit. Another modification of the intrinsic CAR model leading to a proper prior has been presented by Sun et al. (2000).



## 4 Bayesian Inference

In order to account for parameter uncertainty and to allow for an underlying spatial structure we consider the count data regression models discussed in Section 2 in a Bayesian context. MCMC will be used for parameter estimation. For more information on Bayesian data analysis and MCMC methods see Gilks et al. (1996) and Gelman et al. (2004).

Assume the response variables  $Y_i$ ,  $i = 1, \dots, n$  to be observed at  $J$  regions. Besides the well known Poisson regression model  $Poi(\mu_i)$  we consider the  $NB(r, \mu_i)$ ,  $GP(\mu_i, \lambda)$ ,  $ZIP(p, \mu_i)$  and  $ZIGP(p, \mu_i, \lambda)$  model. In each of these models the parameter  $\mu_i$ ,  $i = 1, \dots, n$  is specified by

$$\mu_i = t_i \exp(\mathbf{x}_i' \boldsymbol{\beta} + \gamma_{R(i)}), \quad (4.1)$$

where  $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ik})$  denotes the vector of covariates and  $t_i$  gives the observation specific exposure which will be treated as an offset. The vector  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)$  denotes the vector of unknown regression parameters. Note, that an intercept  $\beta_0$  is included in the model. To allow for geographical differences in the  $J$  regions spatial random effects  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_J)$  are introduced,  $R(i) \in \{1, \dots, J\}$  denotes the region of the  $i$ -th observation. For the zero inflated models we assume a constant  $p$  for all observations. The parameters  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$ ,  $\lambda$ ,  $p$  and  $r$ , respectively, are taken to be a priori independent and the following prior distributions are chosen:

- $\pi(\boldsymbol{\beta}) \sim N(0, \tau^2 I_{k+1})$ , with  $\tau^2 = 100$
- $\pi(\boldsymbol{\gamma} | \sigma^2, \psi) \sim N(\mathbf{0}, \sigma^2 Q^{-1})$  with  $Q$  specified as in (3.2)

For the hyperparameters  $\sigma^2$  and  $\psi$  the proper priors

- $\pi(\sigma^2) \sim IGamma(a, b)$  with  $a = 1$  and  $b = 0.005$
- $\pi(\psi) \sim \frac{1}{(1+\psi)^2}$

are assumed. For the model specific parameters the following prior distributions are chosen:

- GP Regression:  $\pi(\lambda) \sim U([0, 1])$
- NB Regression :  $\pi(r) \sim Gamma(a, b)$ , i.e.  $\pi(r) = \frac{b^a}{\Gamma(a)} r^{a-1} e^{-rb}$ , where  $a = 1$  and  $\pi(b) \sim Gamma(c, d)$ , i.e.  $\pi(b) \propto b^{c-1} e^{-bd}$  with  $c = 1$  and  $d = 0.005$
- ZIP/ZIGP Regression:  $\pi(p) \sim U([0, 1])$

The schemes of the MCMC algorithms and details about the chosen proposal distributions for Metropolis Hastings steps can be found in the Appendix.

## 5 Bayesian Model comparison

### 5.1 Deviance Information Criterion (DIC)

Spiegelhalter et al. (2002) suggest to use the following criterion for model comparison in Bayesian inference. Assume a probability model  $p(\mathbf{y} | \boldsymbol{\theta})$ . The

Bayesian deviance  $D(\boldsymbol{\theta})$ , which is used as a measure for goodness of fit, is defined as

$$D(\boldsymbol{\theta}) = -2 \log p(\mathbf{y}|\boldsymbol{\theta}) + 2 \log f(\mathbf{y})$$

where  $f(\mathbf{y})$  is some fully specified standardizing term. To measure the model complexity Spiegelhalter et al. (2002) introduce the effective number of parameters  $p_D$  defined by

$$p_D := E[D(\boldsymbol{\theta}|\mathbf{y})] - D(E[\boldsymbol{\theta}|\mathbf{y}]).$$

Finally they define the deviance information criterion (DIC) as the sum of the posterior mean of the deviance and the effective number of parameters

$$DIC := E[D(\boldsymbol{\theta}|\mathbf{y})] + p_D.$$

According to this criterion the model with the smallest DIC is to be preferred.  $p_D$  and  $DIC$  are easily computed using the available MCMC output by taking the posterior mean of the deviance to obtain  $E[D(\boldsymbol{\theta}|\mathbf{y})]$  and the plug-in estimate of the deviance  $D(E[\boldsymbol{\theta}|\mathbf{y}])$  using the posterior means  $E[\boldsymbol{\theta}|\mathbf{y}]$  of the parameter  $\boldsymbol{\theta}$ .

Bayes factors based on marginal likelihood provide an alternative method for model comparison, see Kass and Raftery (1995). Further, Bayesian Model Averaging (BMA), see for example Hoeting et al. (1999), which is based on Bayes factors, presents a method for model selection taking model uncertainty into account. However, since the computation of Bayes factors requires substantial efforts for complex hierarchical models, see Han and Carlin (2001), we use the DIC for model comparison in this paper.

## 5.2 Proper scoring rules

Apart from the DIC we use proper scoring rules for categorical variables for comparing models, in particular we consider the Brier score and the logarithmic score presented for example in Gneiting and Raftery (2004). While Gneiting and Raftery (2004) use scoring rules for assessing the quality of probabilistic forecasts, our focus is model comparison based on the posterior predictive distribution. Under the probability model  $p(\mathbf{y}|\boldsymbol{\theta})$  a scoring rule assigns a value  $S(\mathbf{p}_i, y_i)$  for each observation  $y_i, i = 1, \dots, n$  based on the posterior predictive probability vector  $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{im})$ . Here the component  $p_{ij} := P(y_i = j|\mathbf{y})$  denotes the posterior predictive probability that the  $i$ -th observation takes the value  $j$  which can be estimated by  $\hat{p}_{ij} := \frac{1}{J} \sum_{k=1}^J p(y_i = j|\boldsymbol{\theta}^k)$  where  $\boldsymbol{\theta}^k, k = 1, \dots, J$ , denotes the  $k$ -th MCMC iterate of  $\boldsymbol{\theta}$  after burnin. For computational reasons we set  $p_{im} := 1 - \sum_{k=1}^{m-1} p_{ik}$  where  $m - 1$  gives the highest response value observed in the data. This ensures that the probability vector  $\mathbf{p}_i$  sums up to 1.

Models are then compared based on the mean score given by

$$S(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n S(\mathbf{p}_i, y_i).$$

We consider positive oriented scores here, i.e. the model with the highest mean score is to be preferred. The Brier score first proposed by Brier (1950) is defined by

$$S(\mathbf{p}_i, y_i) = 2p_{iy_i} - 1 - \sum_{k=1}^m p_{ik}^2$$

where  $p_{iy_i} = P(y = y_i | \boldsymbol{\theta})$  denotes the posterior predictive probability for the true value  $y_i$  under the considered model. The Brier score corresponds to the expression

$$-\frac{1}{n} \sum_{k=1}^m \sum_{i=1}^n (p_{ij} - \hat{p}_{ij}^{emp})^2$$

where  $\hat{p}_{ij}^{emp} = \begin{cases} 1 & y_i = j \\ 0 & \text{otherwise} \end{cases}$  denotes the empirical probability that the  $i$ -th observation takes the value  $j$ . Hence, according to the Brier score the model which minimizes the squared difference between the observed and the estimated probabilities is considered best.

The logarithmic score is defined by

$$S(\mathbf{p}_i, y_i) = \log p_{iy_i}$$

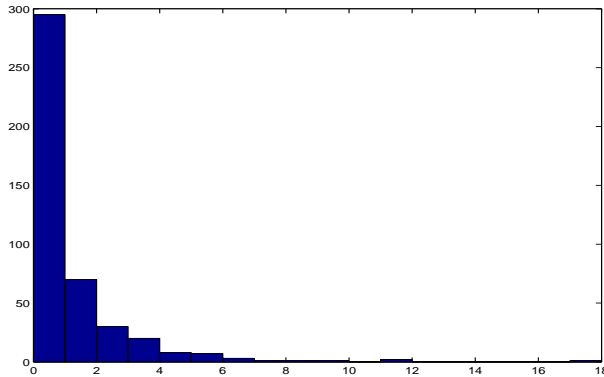
and therefore chooses the model which gives the highest probability for observing the true value. Both scores are proper, i.e. the highest score is obtained for the true model, see Gneiting and Raftery (2004) for details. Further, when parameter estimation is done using MCMC both scores are computed easily based on the available MCMC output as indicated above. Ideally, in order to avoid using the data twice, parameter estimation should be based on about 75 % of the data only, whereas the scores should be computed for the remaining 25 % of the data. However, since the data set analysed in the next section is rather small and our focus is on comparing models rather than prediction, we will use the same data for estimation and computation of the scores.

## 6 Application

### 6.1 Data description

In this section the proposed models will be used to analyse the number of invasive meningococcal disease cases reported in Germany during the year 2004. Meningococcal disease is caused by bacteria and can lead to serious, perilous diseases, like for example meningitis, in which case we refer to invasive meningococcal disease. In 2004, 600 cases of invasive meningococcal disease were reported in Germany. Germany is divided into 439 regions, for each of these regions the number of invasive meningococcal disease cases is given for both men and women. A histogram of the total number of cases in each region is given in Figure 2. A high proportion (67.2 %) of the data is

equal to zero, on average 1.37 cases of meningococcal disease are observed in each region, the maximum number of cases observed in one region is 18. The variance of the data is 3.71 which is substantially higher than the mean and therefore already indicates the presence of overdispersion in the data. On



**Fig. 2** Histogram of the observations  $y_i, i = 1, \dots, 439$ .

a higher aggregation level Germany consists in 16 states. Besides the modelling of overdispersion in the data, an interesting issue is to detect whether there are areas with an increased risk of contracting invasive meningococcal disease. In this case, vaccination could be strongly recommended in these risky regions. Therefore we include the 16 states as covariates in our model, which will be modelled as factor covariates with state 1 as reference level. Since we are interested in relative risks, population effects are eliminated by including the expected number of cases in each region as an offset in the analysis. The expected number of cases in each region is determined by the population in each region times the overall observed risk, i.e. the total number of cases divided by the total population.

Extra heterogeneity in the data, which might not be satisfactorily explained by the gender or the state factors, can be handled by the model specific dispersion parameters in the NB, the GP and the ZI models. While overdispersion in the NB and the GP model can be interpreted as unobserved heterogeneity among observations, zero inflated models would assume that part of the observations equal to zero are extra zeros, i.e. in some regions the occurrence of invasive meningococcal disease might not have been reported. On the other hand, heterogeneity in the data might also be taken into account by assuming a finer geographic resolution, i.e. by including a random spatial effect for each region. We will assume the CAR prior presented in Section 3 for these spatial effects which allows for a spatial dependency structure. In contrast to this approach, the effects of the states included as factor covariates can be seen as unstructured effects on a lower resolution, since no correlation between states is allowed.

We first analysed the data set in Splus using a Poisson model without spatial effects including an intercept and as covariates the gender and the 16 states as factor covariates. No significant influence of gender could be detected, therefore we decided to model the total number of cases without distinguishing between men and women. This means, that we have only one observation  $y_i, i = 1, \dots, 439$  for each region. The four states Nordrhein-Westfalen, Mecklenburg-Vorpommern, Sachsen-Anhalt and Thüringen were found to be significant and only the regression indicators of these states will be included in the following. This model was used as a starting model for the MCMC algorithms of the models discussed in Section 2.

## 6.2 Models

The MCMC algorithms for the Poisson, the GP, the NB, the ZIP and the ZIGP regression models are run for 20000 iterations. The parameter  $\mu_i, i = 1, \dots, n$  is specified for all models by

$$\mu_i = t_i \exp(\mathbf{x}'_i \boldsymbol{\beta} + \gamma_i)$$

with the same covariates included in each model and  $t_i = pop_i \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n pop_i}$  where  $pop_i$  denotes the population in region  $i$ . All models are fitted with and without spatial effects. Since we have only one observation for each region we use the simplified notation  $\gamma_{R(i)} = \gamma_i$ . The first 5000 iterations of the MCMC samplers are discarded as burnin, convergence is achieved well before for all models. After convergence the mixing of the samplers is satisfactorily good, the estimated empirical autocorrelations with lag 5 are in general well below 0.05 for the regression parameters in all non spatial models and both the regression parameters and the spatial effects in the spatial GP and ZIGP models. Only in the spatial Poisson, ZIP and NB models a lag of about 20 is needed in order to obtain autocorrelations of the regression parameters below 0.05, for the spatial effects the autocorrelations are below 0.05 at a lag of 5 in the Poisson and ZIP models and a lag of about 10 in the NB model. The estimated posterior means and 90 % credible intervals for the regression parameters are reported in Table 1 for all models. Estimation of the intercept slightly differs between the models and also changes when spatial effects are added, especially for the Poisson and ZIP models where large spatial effects are observed, see below. Estimation of the state effects is rather similar in all models. For a comparison of the estimated overdispersion in the different models, we consider the estimated dispersion factors  $\varphi_i$  which are defined by  $1 + \frac{\mu_i}{r}, \frac{1}{(1-\lambda)^2}, (p\mu_i + 1)$  and  $p\mu_i + \frac{1}{(1-\lambda)^2}$  for the NB, GP, ZIP and ZIGP regression models, respectively. In particular, we compute the mean, minimum, maximum value and quantiles of the estimated posterior means  $\hat{\varphi}_i := \frac{1}{R} \sum_{j=1}^R \hat{\varphi}_i^j$  of the dispersion factors in each

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
no spatial effects					
Poi	-0.17	0.43	0.65	0.56	0.56
	(-0.26, -0.09)	(0.30, 0.57)	(0.28, 1.02)	(0.23, 0.87)	(0.22, 0.88)
NB	-0.15	0.42	0.67	0.54	0.48
	(-0.25, -0.04)	(0.23, 0.60)	(0.23, 1.10)	(0.16, 0.91)	(0.09, 0.85)
GP	-0.16	0.42	0.65	0.45	0.50
	(-0.26, -0.06)	(0.26, 0.58)	(0.22, 1.06)	(0.05, 0.84)	(0.09, 0.88)
ZIP	-0.13	0.43	0.69	0.60	0.59
	(-0.23, -0.03)	(0.27, 0.59)	(0.29, 1.08)	(0.25, 0.94)	(0.24, 0.95)
ZIGP	-0.15	0.42	0.67	0.47	0.52
	(-0.26, -0.04)	(0.23, 0.61)	(0.22, 1.09)	(0.05, 0.87)	(0.09, 0.92)
with spatial effects					
Poi	-0.25	0.40	0.68	0.58	0.47
	(-0.38, -0.14)	(0.18, 0.61)	(0.21, 1.13)	(0.19, 0.97)	(0.06, 0.86)
NB	-0.18	0.41	0.68	0.55	0.48
	(-0.29, -0.06)	(0.20, 0.61)	(0.23, 1.21)	(0.17, 0.93)	(0.08, 0.87)
GP	-0.16	0.42	0.65	0.45	0.50
	(-0.26, -0.06)	(0.25, 0.58)	(0.21, 1.06)	(0.04, 0.84)	(0.09, 0.88)
ZIP	-0.23	0.39	0.68	0.60	0.49
	(-0.36, -0.10)	(0.16, 0.63)	(0.22, 1.14)	(0.19, 0.99)	(0.07, 0.90)
ZIGP	-0.15	0.42	0.66	0.48	0.52
	(-0.26, -0.04)	(0.23, 0.61)	(0.22, 1.09)	(0.06, 0.88)	(0.10, 0.92)

**Table 1** Posterior means and 90 % credible intervals for the regression parameters ( $\beta_1, \dots, \beta_4$ : effects for the states Nordrhein-Westfalen, Mecklenburg-Vorpommern, Sachsen-Anhalt and Thüringen, respectively) in the different models for the meningococcal disease data.

model, where  $\hat{\varphi}_i^j$  denotes the  $j$ -th MCMC iterate for  $\varphi_i$  after burnin. The results are reported in Table 2. Note, that the dispersion factor in the GP regression model is the same for all observations, whereas it depends on the parameter  $\mu_i$  and therefore is different for each observation in the other models. Except for the ZIP model, all models exhibit a substantial degree of overdispersion with respect to the Poisson model, regardless whether spatial effects are included or not. In the non spatial NB model the average of the estimated posterior means of the dispersion parameter is given by 1.396 and drops to 1.293 when spatial effects are included. The range of the estimated spatial effects in the NB model, see Table 3, is considerably smaller than in the Poisson model where unexplained heterogeneity in the data is captured by the spatial effects alone. However, part of the data variability in the NB model is explained by spatial effects as well rather than the parameter  $r$  alone. This is in contrast to the GP and ZIGP model, where the estimated spatial effects are all very close to zero. Overdispersion in these models is captured by the parameter  $\lambda$  only, resulting in a high estimated dispersion parameter. Results are hardly affected by the inclusion of spatial effects. The extension from a GP to a ZIGP model has almost no influence on the

Parameter	$\gamma$	$\hat{\varphi}_i$					
		mean (2.5%,97.5 %)	min	25%	50%	75%	max
$r$ in NB	yes	6.433 (2.483, 22.552)	1.051	1.150	1.293 1.205	1.313	5.358
	no	3.836 (2.202, 6.864)	1.064	1.200	1.396 1.273	1.428	7.053
$\lambda$ in GP	yes	0.162 (0.098, 0.231)	1.432				
	no	0.163 (0.098, 0.232)	1.435				
$p$ in ZIP	yes	0.029 (0.001, 0.087)	1.007	1.020	1.041 1.028	1.044	1.536
	no	0.056 (0.006, 0.125)	1.013	1.041	1.081 1.056	1.089	2.226
$p$ in ZIGP	yes	0.019 (0.001, 0.064)	1.434				
$\lambda$ in ZIGP	yes	0.155 (0.088, 0.224)	1.412	1.421	1.426	1.436	1.819
$p$ in ZIGP	no	0.019 (0.001, 0.063)	1.440				
$\lambda$ in ZIGP	no	0.157 (0.090, 0.227)	1.417	1.427	1.432	1.442	1.826

**Table 2** Estimated posterior means for the model specific dispersion parameters in the considered models with and without spatial effects, with the 2.5 % and 97.5 % quantiles given in brackets. Further the mean (upper row), range and quantiles (lower row) of the estimated posterior means of the dispersion factors  $\hat{\varphi}_i$  are given.

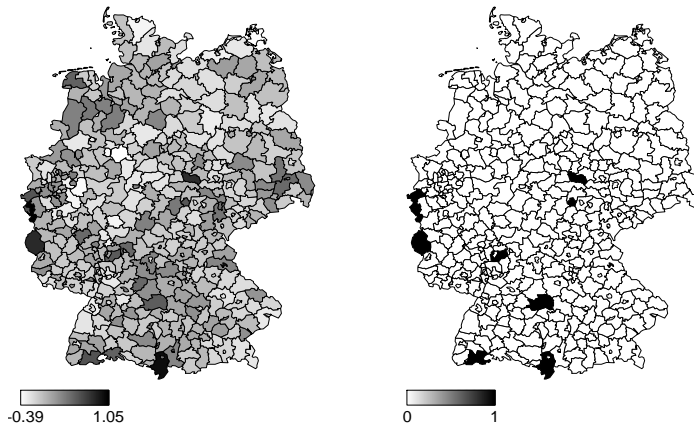
estimation of  $\lambda$  and the average dispersion parameter  $\varphi_i$ , the proportion of extra zeros  $p$  is estimated very close to zero.

In the non spatial ZIP model the proportion of extras zeros  $p$  is estimated as 5.6 %, resulting in an average dispersion parameter of about 1.081. According to the large 95 % credible interval for  $p$  however, no significant degree of zero inflation seems to be present. Unobserved heterogeneity still present in the data after adjusting for covariates is captured better by the GP and NB model, whereas the assumptions of extra zeros is obviously not appropriate for this data. When spatial effects are included to the ZIP model the estimated proportion of extra zeros drops even further, indicating that unexplained heterogeneity is picked up mostly by the spatial effects alone like in the Poisson model, the range of the estimated posterior means of the spatial effects in the ZIP and the Poisson model is almost the same, see Table 3. The map plot of the estimated posterior means of the spatial effects in the Poisson model, given in Figure 3 roughly represents the spatial pattern of the observed relative risk in each region  $\frac{y_i}{t_i}, i = 1, \dots, n$  which is plotted in the upper row in Figure 4. The estimated posterior mean and median of the spatial hyperparameter  $\psi$  in the Poisson model, see Table 3

Model	$[\min_j \hat{\gamma}_j \max_j \hat{\gamma}_j]$	$\psi$		$\sigma^2$	
		mean	median	mean	median
Poisson	$[-0.383, 1.059]$	0.394 (0.011,0.431)	0.207	0.541 (0.149, 0.615)	0.422
NB	$[-0.151, 0.345]$	1.223 (0.015,1.352)	0.588	0.227 (0.012,0.281)	0.140
GP	$[-0.009, 0.020]$	2.604 (0.038,2.199)	1.194	0.012 (0.001,0.011)	0.005
ZIP	$[-0.371, 0.998]$	0.329 (0.013,0.397)	0.235	0.477 (0.169,0.529)	0.417
ZIGP	$[-0.011, 0.019]$	1.739 (0.044,1.905)	1.154	0.018 (0.001,0.027)	0.011

**Table 3** Range of estimated posterior means of the spatial effects as well as estimated posterior means, medians and 95 % credible intervals given in brackets for the spatial hyperparameters in the considered models for the meningococcal disease data.

are rather small, the lower bound of the 95 % credible interval is close to zero, indicating that the overall degree of spatial dependence is very small. This is reflected in the estimated spatial pattern which is not particularly smooth. Only some rather isolated regions, which are marked in black in

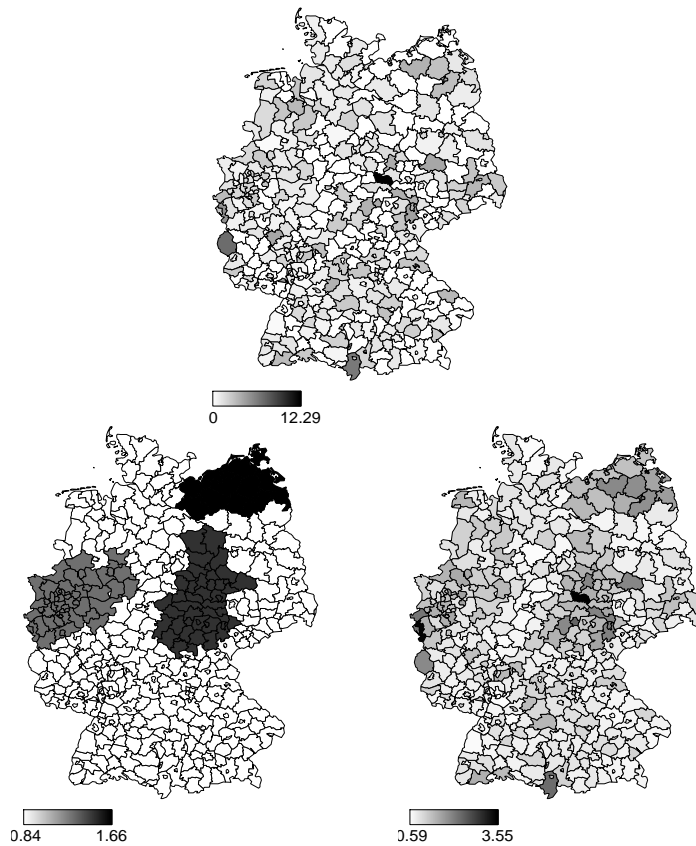


**Fig. 3** Maps of the estimated posterior means (left) and 80 % credible intervals (white: 0 included in credible interval, black: strictly negative credible interval) of the spatial effects in the Poisson regression model for the Meningococcal disease data.

the right map in Figure 3, have a significant positive spatial effect according to the 80 % credible intervals. In these regions the observed number of inva-



sive meningococcal disease cases was rather high and most of them do not lie within the four states included as covariates. Therefore without spatial effects the risk in these regions is not modelled sufficiently. The estimated



**Fig. 4** Maps of the observed relative risk  $\frac{\mu_i}{t_i}$  (upper row) and the estimated posterior means of the risk factor  $\frac{\mu_i}{t_i}$  in the non-spatial (lower row, left) and spatial (lower row, right) Poisson regression model for the meningococcal disease data.

posterior means of the risk factor  $\frac{\mu_i}{t_i}$  are plotted for the non-spatial and spatial Poisson model in the lower row of Figure 4. In the non-spatial Poisson model geographic differences are modelled by four state indicators only. Since the risk in two of the states, which are neighbours, is about the same, visually only three states can be distinguished in this plot. The inclusion of spatial random effects gives a rather smoothed representation of the true pattern, however the rough structure is detected reasonable well.

### 6.3 Model comparison using DIC

In order to compare the presented models the DIC, reviewed in Section 5.1, is considered. In Table 4, the DIC, the posterior mean of the deviance and the effective number of parameters are given for each model. Only in the Poisson regression case a well defined normalizing constant  $f(y)$  (see Section 5.1) exists, while in all other models the likelihood of the saturated model depends on the unknown overdispersion parameters. Therefore we make the choice of setting the normalizing function  $f(y)$  to 0. Consequently  $E[D(\boldsymbol{\theta}|\mathbf{y})]$  is based only on the unscaled deviance which cannot be directly interpreted as an overall goodness of fit measure of one specific model. However,  $E[D(\boldsymbol{\theta}|\mathbf{y})]$  can be used for comparing the model fit of several models when the number of parameters is roughly the same.

Model	$\gamma$	DIC	$E[D(\boldsymbol{\theta} \mathbf{y})]$	$p_D$
Poisson	no	1291.8	1286.8	5.04
NB	no	1273.9	1267.8	6.10
GP	no	1265.6	1259.6	6.01
ZIP	no	1291.8	1285.9	5.96
ZIGP	no	1267.8	1261.5	6.35
Poisson	yes	1248.7	1159.1	89.56
NB	yes	1270.8	1240.0	30.74
GP	yes	1265.7	1258.3	7.32
ZIP	yes	1255.4	1175.1	80.31
ZIGP	yes	1267.6	1260.2	8.28

**Table 4** DIC,  $E[D(\boldsymbol{\theta}|\mathbf{y})]$  and effective number of parameters  $p_D$  for the different models.

For the non spatial models the lowest value of the DIC is obtained for the GP model, while the DIC for the Poisson and the ZIP model takes the highest value. Hence, according to the DIC the GP model is considered best among the non spatial models, while the Poisson and ZIP model clearly perform worse. The effective number of parameters  $p_D$  is close to the true number of parameters which is five for the Poisson regression model, six for the NB, GP and ZIP regression models and seven for the ZIGP regression model.

When spatial effects are added, the posterior mean of the deviance and the number of effective parameters in the GP and ZIGP models hardly change. As mentioned in the previous section already, spatial effects are not significant in these models, i.e. after adjusting for covariate information, there is no further spatial heterogeneity in the data which might be captured by the spatial effects. Instead any overdispersion present in the data seems to be sufficiently captured by the model specific dispersion parameter. The DIC for the spatial NB model is slightly better than for the non spatial one, hence spatial effects improve the model. However, the spatial pattern

is rather smooth as can be seen from the effective number of parameters estimated by 30.74. For the Poisson and ZIP regression model in contrast, a significant drop in the DIC is observed when spatial effects are taken into account. This shows that there is some extra variability in the data which is not sufficiently explained by the covariates only in these models. Since the Poisson model does not allow for overdispersion and the heterogeneity is not of a zero inflated nature, for these two models the unexplained variability is covered by the spatial effects. According to the DIC the spatial Poisson model gives the best fit and is to be preferred to a non spatial GP model. Note, that the DIC must be used with care here, since strictly speaking the DIC is defined for distributions of the exponential family only, see van der Linde (2005). However, the posterior mean of the deviance  $E[D(\boldsymbol{\theta}|\mathbf{y})]$  which can be considered for comparing the model fit of the non spatial models where the number of parameters is very close, gives the same ranking of the models as the DIC.

#### 6.4 Model checking using proper scoring rules

Apart from the DIC we also compute the Brier score and the logarithmic score presented in Section 5.2 for each model, results are reported in Table 5. These scores are based both on the posterior predictive probabilities and the true observed number of cases and therefore provide a good measure for checking which model fits the data best. The results support the conclusions

Model	$\gamma$	Brier score	logarithmic score
Poisson	no	-0.6937	-1.4569
NB	no	-0.6883	-1.4363
GP	no	-0.6873	-1.4272
ZIP	no	-0.6921	-1.4549
ZIGP	no	-0.6878	-1.4291
Poisson	yes	-0.6280	-1.2422
NB	yes	-0.6717	-1.3779
GP	yes	-0.6863	-1.4243
ZIP	yes	-0.6481	-1.3251
ZIGP	yes	-0.6900	-1.4529

**Table 5** Brier score and logarithmic score for the considered models with and without spatial effects.

drawn in the previous section. For the non spatial models the GP regression model fits the heterogeneity in the data best, followed by the ZIGP and NB regression model. The use of a non spatial ZIP regression model does not seem to be appropriate, the gain in comparison to the non spatial Poisson model for which the lowest scores are obtained is very small. The scores for the GP and the ZIGP model hardly change by allowing for spatial effects,

indicating that the model specific dispersion parameters capture the data heterogeneity well. Again a small improvement in the NB model is observed when spatial effects are included. The scores for the spatial Poisson and ZIP model however, are considerably smaller than for the other models. This confirms again, that spatial effects have a significant influence in these models and that a spatial Poisson model gives the best fit to the data.

## 7 Conclusions

We have presented several regression models for count data allowing for overdispersion. Overdispersion is either modelled by the introduction of an additional parameter as in the NB and GP model, by allowing for an extra proportion of zero observations using zero inflated models or by combining zero inflated models with overdispersed distributions.

Further, additionally spatial random effects are included in the models in order to account for unobserved spatial heterogeneity in the data. This approach allows for spatial correlations between observations.

These models were applied to analyse the number of invasive meningococcal disease cases in Germany in the year 2004. The DIC, the Brier and the logarithmic score were used for model comparison. The models allowing for overdispersion gave a significantly better fit than an ordinary non spatial Poisson regression model. Among these non spatial models, the GP model fitted the data best, while the overdispersion present did not seem to be caused by the presence of extra zeros in the data. For the GP and the ZIGP model the inclusion of spatial effects did not improve the models, in the NB model still some significant spatial variation was detected. For the Poisson model which does not allow for overdispersion and the ZIP model which is not modelling the nature of the overdispersion appropriately, the inclusion of spatial effects led to a significant improvement. According to the considered criterions the spatial Poisson model is to be preferred to all other models. But we would like to note that the spatial model fitted shows no smooth surface structure, it rather indicates isolated specific regions where the covariates provide no adequate fit.

Instead of analysing the number of cases of invasive meningococcal disease for one year only, it might be interesting to include data over several years in the analysis. Space-time interactions could be included in order to examine whether the spatial pattern changed over the years. This is the topic of future research.

## A Appendix

In the following the algorithmic schemes of the MCMC samplers for the discussed models are summarized. Most update steps are performed using

a single component Metropolis Hastings (MH) step. For the proposal distributions either a symmetric random walk proposal or an independence proposal is used. In particular, for the independence proposal we take a t-distribution with  $v = 20$  degrees of freedom with the same mode and the same inverse curvature at the mode as the target distribution.

### A.1 GP regression model

- Sample  $\lambda | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}$
- Sample  $\beta_j | \mathbf{y}, \lambda, \boldsymbol{\beta}_{-j}, \boldsymbol{\gamma}, \quad j = 0, \dots, k$
- Update of spatial effects
  - Sample  $\frac{1}{\sigma^2} | \boldsymbol{\gamma}, \psi \sim \text{Gamma}$
  - Sample  $\psi | \boldsymbol{\gamma}, \sigma$
  - Sample  $\gamma_j | \mathbf{y}, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}_{-j}, \psi, \sigma, \quad j = 1, \dots, J$

Since the full conditional of  $\sigma^2$  is Inverse Gamma,  $\sigma^2$  can be sampled directly using a Gibbs step. For the remaining parameters a MH step is used. In particular,  $\lambda$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  are updated component by component using an independence proposal distribution. The spatial hyperparameter  $\psi$  is updated using a random walk proposal. For the Poisson regression model the algorithmic scheme is the same, but with  $\lambda$  set fix to 0.

### A.2 NB regression model

- Sample  $r | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}$
- Sample  $\beta_j | \mathbf{y}, r, \boldsymbol{\beta}_{-j}, \boldsymbol{\gamma}, \quad j = 0, \dots, k$
- Update of spatial effects
  - sample spatial hyperparameters  $\frac{1}{\sigma^2}$  and  $\psi$  as in A.1
  - Sample  $\gamma_j | \mathbf{y}, r, \boldsymbol{\beta}, \boldsymbol{\gamma}_{-j}, \psi, \sigma, \quad j = 1, \dots, J$

In the NB regression model  $r$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  are updated component by component using a MH step. For  $r$  a random walk proposal is used, while  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}$  are updated using an independence sampler.

### A.3 ZI models

To avoid convergence problems in the ZI models which arose in simulated data due to correlation between the intercept  $\beta_0$ ,  $p$  and  $\lambda$ , we use collapsed algorithms, in particular  $\beta_0$ ,  $p$  and  $\lambda$  are updated with the latent variables  $\mathbf{z}$  integrated out, i.e. based on model (2.5). Doing so convergence and mixing of the samplers was improved a lot.

### A.3.1 ZIP model with constant $p$

- Updates with  $\mathbf{z}$  integrated out
  - Sample  $\beta_0 | \mathbf{y}, p, \beta_{-0}, \gamma$
  - Sample  $p | \mathbf{y}, \beta, \gamma$
- Sample  $z_i | \mathbf{y}, p, \beta, \gamma \sim \text{Bernoulli}\left(\frac{p}{p + (1-p)\exp(-\mu_i)}\right) \quad \forall i \text{ with } y_i = 0$
- Sample  $\beta_j | \mathbf{y}, \beta_{-j}, \mathbf{z}, \gamma, \quad j = 1, \dots, k$
- Update of spatial effects
  - sample spatial hyperparameters  $\frac{1}{\sigma^2}$  and  $\psi$  as in A.1
  - Sample  $\gamma_j | \mathbf{y}, \beta, \mathbf{z}, \gamma_{-j}, \psi, \sigma, \quad j = 1, \dots, J$

The latent variables  $\mathbf{z}$  can be updated using a Gibbs step. Since the full conditional of  $p$  is log concave, adaptive rejection sampling (ARS) introduced by Gilks and Wild (1992) is used to update  $p$ . For the parameters  $\beta$  and  $\gamma$  a MH step using an independence proposal distribution is performed.

### A.3.2 ZIGP model with constant $p$

- Updates with  $\mathbf{z}$  integrated out
  - Sample  $\beta_0 | \mathbf{y}, p, \lambda, \beta_{-0}, \gamma$
  - Sample  $p | \mathbf{y}, \lambda, \beta, \gamma$
  - Sample  $\lambda | \mathbf{y}, p, \beta, \gamma$
- Sample  $z_i | \mathbf{y}, \lambda, p, \beta, \gamma \sim \text{Bernoulli}\left(\frac{p}{p + (1-p)\exp(-\mu_i)}\right) \quad \forall i \text{ with } y_i > 0$
- Sample  $\beta_j | \mathbf{y}, \lambda, \beta_{-j}, \mathbf{z}, \gamma, \quad j = 1, \dots, k$
- Update of spatial effects
  - sample spatial hyperparameters  $\frac{1}{\sigma^2}$  and  $\psi$  as in A.1
  - Sample  $\gamma_j | \mathbf{y}, \lambda, \beta, \mathbf{z}, \gamma_{-j}, \psi, \sigma, \quad j = 1, \dots, J$

For the ZIGP model the same proposal distributions as in the ZIP model are used. For  $\lambda$  an independence proposal is taken.

## Acknowledgement

The first author is supported by a doctoral fellowship within the Graduiertenkolleg *Angewandte Algorithmische Mathematik*, while the second author is supported by Sonderforschungsbereich 386 *Statistische Analyse Diskreter Strukturen*, both sponsored by the *Deutsche Forschungsgemeinschaft*.

## References

- Agarwal, D. K., A. E. Gelfand, and S. Citron-Pousty (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics* 9, 341–355.

- Angers, J.-F. and A. Biswas (2003). A Bayesian analysis of zero-inflated generalized Poisson model. *Computational Statistics & Data Analysis* 42, 37–46.
- Banerjee, S., B. Carlin, and A. Gelfand (2004). *Hierarchical Modeling and Analysis for Spatial Data*. New York: Chapman & Hall/CRC.
- Besag, J. and C. Kooperberg (1995). On conditional and intrinsic autoregressions. *Biometrika* 82, 733–746.
- Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78 (1), 1–3.
- Consul, P. (1989). *Generalized Poisson Distributions. Properties and Applications*. New York: Marcel Dekker, Inc.
- Consul, P. and G. Jain (1973). A generalization of the Poisson distribution. *Technometrics* 15, 791–799.
- Czado, C. and S. Prokopenko (2004). Modeling transport mode decisions using hierarchical binary spatial regression models with cluster effects. *Discussion paper 406, SFB 386 Statistische Analyse diskreter Strukturen*. <http://www.stat.uni-muenchen.de/sfb386/>.
- Famoye, F. and K. Singh (2003a). On inflated generalized Poisson regression models. *Advances and Applications in Statistics* 3 (2), 145–158.
- Famoye, F. and K. Singh (2003b). Zero inflated generalized Poisson regression model. *submitted*.
- Gelman, A., J. Carlin, H. Stern, and D. Rubin (2004). *Bayesian Data Analysis, Second Edition*. Boca Raton: Chapman & Hall/CRC.
- Gilks, W., S. Richardson, and D. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. Boca Raton: Chapman & Hall/CRC.
- Gilks, W. and P. Wild (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 41 (2), 337–348.
- Gneiting, T. and A. E. Raftery (2004). Strictly proper scoring rules, prediction and estimation. Technical report no. 463, Department of Statistics, University of Washington.
- Han, C. and B. Carlin (2001). Markov chain Monte Carlo methods for computing bayes factors: A comparative review. *Journal of the American Statistical Association* 96, 1122–1132.
- Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science* 14 (4), 382–417.
- Jin, X., B. Carlin, and S. Banerjee (2004). Generalized hierarchical multivariate CAR models for areal data. *submitted to Biometrics*.
- Joe, H. and R. Zhu (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with Negative Binomial distribution. *Biometrical Journal* 47, 219–229.
- Kass, R. and A. Raftery (1995). Bayes factors and model uncertainty. *Journal of the American Statistical Association* 90, 773–795.
- Lambert, D. (1992). Zero-inflated Poisson regression with and application to defects in manufacturing. *Technometrics* 34 (1), 1–14.

- Pettitt, A., I. Weir, and A. Hart (2002). A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing* 12 (4), 353–367.
- Rodrigues, J. (2003). Bayesian analysis of zero-inflated distributions. *Communications in Statistics* 32 (2), 281–289.
- Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B* 64 (4), 583–640.
- Sun, D., R. K. Tsutakawa, H. Kim, and Z. He (2000). Bayesian analysis of mortality rates with disease maps. *Statistics in Medicine* 19, 2015–2035.
- van der Linde, A. (2005). DIC in variable selection. *Statistica Neerlandica* 59 (1), 45–56.
- Winkelmann, R. (2003). *Econometric Analysis of Count Data. 4th Edition*. Berlin: Springer-Verlag.