DEPT. OF MATH. UNIVERSITY OF OSLO
STATISTICAL RESEARCH REPORT NO 2
ISSN 0806–3842 JUNE 2010

A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims. Derivation of distributions and MCMC sampling schemes

Ida Scheel* † ‡, Egil Ferkingstad§ ‡, Arnoldo Frigessi¶ ‡,
Ola Haug§ ‡, Mikkel Hinnerichsen‖, Elisabeth Meze-Hausken‖

Abstract

Climate change will affect the insurance industry. We develop a Bayesian hierarchical statistical approach to explain and predict insurance losses due to weather events at a local geographical scale. The number of weather-related insurance claims is modelled combining generalized linear models with spatially smoothed variable selection. Using Gibbs sampling and reversible jump MCMC, the model is fitted on daily weather and insurance data from each of the 319 municipalities of southern and central Norway for the period 1997-2006. Out-of-sample predictions from the model are very good. Our results show interesting regional patterns in the impact of different weather covariates. In addition to being useful for insurance pricing, our model can be used for short-term predictions based on weather forecasts and long-term predictions based on downscaled climate models.

Keywords: Bayesian Poisson Hurdle, Zero-Altered Poisson, Spatial Variable Selection, Hierarchical Models, Generalized Linear Models, Climate Change.

1 Introduction

The global insurance industry is highly exposed to risks caused by weather related events. In the past two decades, there are clear signs of a significant increase in the number of claims, possibly due to changes in the spatial distribution, frequency, and intensity of both ordinary and catastrophic weather events. Simultaneously, demographic and socio-economic trends are increasing society's exposure to weather-related losses. An analysis of insurance industry data showed that weather-related catastrophic losses (hurricanes, storms, floods, extreme draughts) have increased by 2% each year since the 1970s, adjusting for changes in wealth, inflation and population growth (Muir-Wood *et al.*, 2006). While extreme and large scale catastrophic events represent roughly 40% of the insured weather-related losses globally, small scale weather related events (such as rain, hailstorms, heavy wind, frost) account for most of the incurred losses (Mills *et al.*, 2005; Botzen & van den Bergh, 2008). Losses due to non-catastrophic weather patterns are expected to increase non-linearly with precipitation intensity. Neglecting this could lead to underestimating losses; this would represent a serious danger to insurance companies.

In order to understand patterns of risk over time and geography, the first step is to explore the relationship between weather events and incurred losses by analysing historical data. In this

*Corresponding author. Email: idasch@math.uio.no. Postal address: Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, 0316 Oslo, Norway
†Department of Mathematics, University of Oslo, Oslo, Norway
‡(sfi)^2- Statistics for Innovation
§Norwegian Computing Center, Oslo, Norway
¶Department of Biostatistics, University of Oslo
‖Gjensidige, Oslo, Norway

paper we analyse the relation between the number of claims and the weather, using ten years of historical insurance and meteorological data in Norway. We focus on damages caused to privately owned buildings, and exclude a small number of catastrophic weather related events, which in Norway are covered by a separate national fund. We consider only the number of claims. Modelling the size of the claims is more simple, but requires adjustment for several economical trends (Frees & Valdez, 2008).

The purpose of our study is (i) to understand which weather patterns are responsible for claims and (ii) to predict the number of losses given a certain weather pattern. In both cases we work at a local scale. Our results can be used to develop strategies to limit the effects of weather events, through preventive actions together with insured customers and local authorities. Also, they help update risk estimation and premium calculations. Our model can be used to predict the number of claims at a regional scale, while it rains, using actually occurring meteorological conditions. This is useful if the actual reporting is delayed. Also, we can provide short term predictions of insurance losses based on weather forecasts one or two weeks ahead. The insurance company can use these predictions to organise inspections and support in the right place immediately for clients whose buildings are damaged.

We can also use the model to investigate the effect of hypothetical weather scenarios, constructed for example by resampling historical events and placing them at different geographical locations. More interestingly, one can use downscaled climate models, to understand the potential exposure of the insured portfolio under future climate conditions.

We cast the problem in a Bayesian space-time setting, where appropriate regressions are performed in each municipality, with weather variables as covariates. We wish to select the relevant covariates locally, but assume that models vary smoothly over space. This leads to the formulation of a hierarchical Bayesian spatial variable selection process. Insurance count data, as we analyse in this paper, show an excess of zeros (many days without claims). Different weather patterns are thought to be responsible for having a day with no claims as opposed to one or more claims, and for the actual count on days with claims. This leads to the Hurdle Poisson model, as in Mullahy (1986), rather than a Zero Inflated Poisson model (Lambert, 1992). We use conjugate Gaussian g-priors for the GLM of the Positive Poisson part (which allows us to integrate out the regression parameters), and a spatial Ising model which steers the variable selection in a spatially smooth way. The local model varies smoothly over the geography of Norway, as we do not expect abrupt changes in the weather-related claims in areas with comparable geographical conditions, and we can therefore borrow strength due to inhomogeneous exposures. Inference can be conveniently split in two separate tasks: one for the Bernoulli probability of a claim, the other for the intensity of the Positive Poisson count. We implemented a Gibbs Sampler for the Positive Poisson intensity and a reversible jump MCMC scheme for the Bernoulli component of the model. When predicting the number of claims, the two components cannot be treated separately any more, and we used predictive Gibbs sampling for the variable dimension parameter space.

Our results show interesting regional patterns of weather covariates contributing to presence/absence of claims and to the number of claims. While precipitation and drainage are important for conditions that lead to at least one claim: the number of claims (i.e. the seriousness of the local damage) seems to be modulated by more complex weather events. For the Positive Poisson intensity, the model with largest posterior probability includes more covariates, and show less smoothness. Snow variables are present in some municipalities. Out-of-sample predictions are very good: on average we predict the right quantity of claims per week in a municipality in 89% of the weeks, and in all but one of the municipalities we predict correctly in more than 50% of the weeks. Our model copes reasonably well with extreme precipitation, but is less able to predict extreme numbers of claims, which happen unrelated to extreme weather.

There has been some research on the relation between weather and the insurance industry (Vellinga *et al.*, 2001; Nordhaus, 2008; Association of British Insurers, 2005), but this area lacks public data, due to their presumed competitive value, and studies are scarce. Some data aggregated in space and time are available, and Mills (2005) identifies the financial services and asset management companies as vulnerable to climate change. Botzen & van den Bergh (2008) discuss various types of risk exposures to weather related events in the insurance industry, concentrating on the case of the Netherlands. We support these authors' call for the industry to share their data. The study presented in this paper shows that insight can be gained by a thorough statistical analysis. Adaptation to climate change for the insurance industry is discussed in Warner *et al.* (2009) and Kleindorfer (2010). Actuarial applications of hierarchical insurance claim models, including several Zero Inflated stochastic models, are presented in the fundamental papers Frees & Valdez (2008); Frees *et al.* (2009); Boucher & Denuit (2006); Boucher & Guillen (2009).

The outline of this paper is as follows: In section 2 we describe our data set. Section 3 contains the hierarchical model. We present results in Section 4, and conclude with a brief discussion in Section 5.

2 Insurance and weather data

The insurance data are from Gjensidige (`www.gjensidige.no`), the largest non-life insurance company in Norway, and include all insured private buildings in the period 1997-2006. For a more complete description of the data see Haug *et al.* (2008, 2009). In this paper we focus on the municipalities of central and south Norway, which have the majority of the claims. For each of the $K = 319$ municipalities, we obtained the daily number of claims due to damages caused by either precipitation, surface water, snow melting, undermined drainage, sewage backflow or blocked pipes. In addition we have the monthly number of insured buildings, for each municipality, representing exposure. For municipality $k, k = 1, 2, ..., K$, N_{kt} is the observed number of claims in day t, T_k is the set of days for which we have observed N_{kt} (as there are a few missing values in the data), t_k is the number of days in T_k. Let $\boldsymbol{N}_k = (N_{k1}, N_{k2}, ..., N_{kt_k})^T$ and $\boldsymbol{A}_k = (A_{k1}, A_{k2}, ..., A_{kt_k})^T$ be the vectors of claims and insured units in municipality k for each time point.

Figure 1 describes the spatial variability of the exposure: for each municipality we computed the average daily number of policies, and plotted the percentage with respect to the maximum exposure (which is in Bergen). Most claims are concentrated on the main cities. The mean claim size in the various municipalities ranged from 20000 NOK to 65000 NOK (price index adjusted, data not shown).

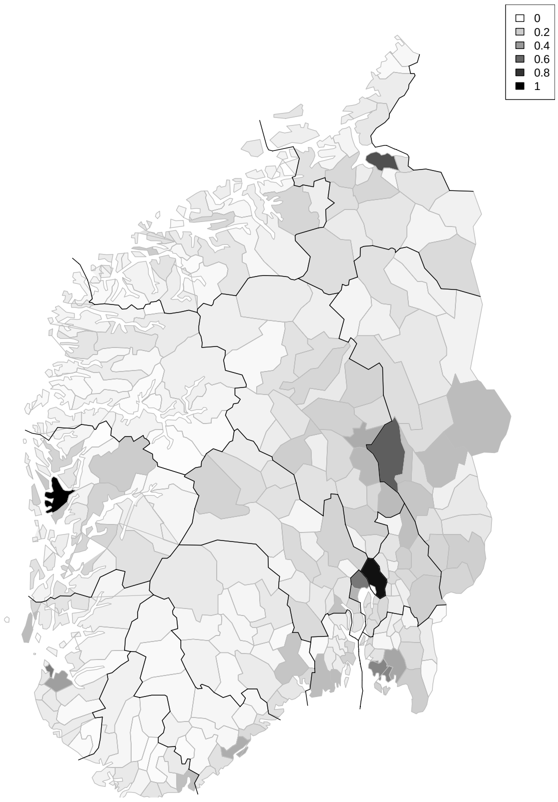
The Norwegian Meteorological Institute (`www.met.no`), together with the Norwegian Water Resources and Energy Directorate (`www.nve.no`), produced meteorological and hydrological data: daily mean precipitation, mean temperature, drainage run-off and snow water equivalent for each municipality, on each day for the period 1997-2006. Data were interpolated from a grid of monitoring stations, weighting areas within each municipality proportionally to the population density. In this way the meteorological covariates describe more accurately the occurred weather at locations of insured buildings for larger municipalities.

Table 1 shows the $q = 7$ covariates we use in our models, four are basic and three derived. The measurement period for precipitation is delayed by 6 hours compared to calendar time. That is, the total precipitation registered in day $t + 1$ is the amount collected from six in the morning of day t to six in the morning of day $t + 1$. R_{3t} is the accumulated rain in the last three days. The difference in snow water equivalent for successive days, S_Δ, can also be relevant. Define $\boldsymbol{X}_k$ as the $t_k \times q$ weather covariate matrix for municipality k. We will use the full covariate matrix as well as a reduced version corresponding to the days with positive number of claims, both of which are appropriately centred and scaled.

3 Bayesian Poisson Hurdle model with Ising smoothed variable selection

The daily number of claims N_{kt} in a municipality k is zero more often than would be modelled with a Poisson distribution. Also, we expect a threshold effect of the weather covariates on claims, as there is no damage caused by normal weather states. Therefore, the number of claims are assumed to follow a Hurdle model (Mullahy, 1986), also known as the two-part model (Heilbron, 1994). The model consists of two parts, one of which is a Bernoulli distribution modelling whether the count is zero or positive. The second part models strictly positive counts by a count distribution.

Figure 1: The exposure for each municipality in central and south Norway described by the average daily number of insurance policies as percentage w.r.t. maximum (Bergen).



Let α_{kt} be the probability of a zero count. The latent binary variable ζ_{kt} indicates whether there is a zero ($\zeta_{kt} = 0$) or positive ($\zeta_{kt} = 1$) count, with an a priori Bernoulli$(1-\alpha_{kt})$ probability. The second component of the Hurdle model, modelling positive counts, is assumed to be Positive Poisson with parameter λ_{kt}. We call the whole model the Bernoulli Poisson Hurdle (BPH) model. The BPH model for the number of claims is hence

$$P(N_{kt} = n \mid \alpha_{kt}, \lambda_{kt}) = \alpha_{kt} \cdot \mathbb{1}_{n=0} + (1-\alpha_{kt}) \frac{\lambda_{kt}^n}{(\exp(\lambda_{kt}) - 1)n!} \cdot \mathbb{1}_{n>0}, \quad k = 1, \ldots K, \ t \in T_k \tag{1}$$

where $\mathbb{1}_C$ equals 1 when C is true, and 0 otherwise, and we write $\boldsymbol{N}_k \mid \boldsymbol{\alpha}_k, \boldsymbol{\lambda}_k \sim BPH(\boldsymbol{\alpha}_k, \boldsymbol{\lambda}_k)$, where $\boldsymbol{\alpha}_k$ is the vector of α_{kt} for all t, and $\boldsymbol{\lambda}_k$ is the vector of λ_{kt} defined only on days with positive count. We model α_{kt} and λ_{kt} by generalized linear models (GLM), separately for each municipality, but with spatially smoothed variable selection. We use a logit-link for α_{kt} in the Bernoulli distribution, and a log-linear model for λ_{kt} in the Positive Poisson distribution, with Gaussian overdispersion. Each municipality has a pair of GLMs for α_{kt} and λ_{kt}. A consequence of the Hurdle model formulation is that λ_{kt} only matters for a day t and municipality k if there is a positive count. Consequently λ_{kt} is dependent on ζ_{kt} and α_{kt}. However, posterior inference for the model separates into two: the zero count part and the positive count part. These can be executed completely separately, as they are conditionally independent given the data. We will first describe the GLM model for λ_{kt}, and then the GLM model for α_{kt}.

In order to model the selection of the covariates appropriate for the GLM for $\boldsymbol{\lambda}_k$ for municipality k, we introduce the vector of binary variables $\boldsymbol{\gamma}_{k\cdot}^{\lambda} = (\gamma_{k1}^{\lambda}, \ldots, \gamma_{kq}^{\lambda})^T$. For municipality k and covariate j, $\gamma_{kj}^{\lambda} = 1$ means that covariate j enters the model for $\boldsymbol{\lambda}_k$, $\gamma_{kj}^{\lambda} = 0$ otherwise. Let $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kq})^T$ be the coefficients of the covariates for the GLM for λ_{kt}, with $\beta_{kj} = 0$ if $\gamma_{kj}^{\lambda} = 0$. Define the reduced vector $\boldsymbol{\beta}_k^{\gamma}$ as the vector of β_{kj} for which $\gamma_{kj}^{\lambda} = 1$. The intercept β_{k0} is always part of the model. We reduce $\boldsymbol{X}_k$ to include only the rows corresponding to the days with positive count, and for convenience denote it still as $\boldsymbol{X}_k$, with a slight abuse of notation.

Variable	Description	Unit
R_t	Precipitation registered day t (mainly collected during day $t-1$)	mm
C_t	Mean temperature in day t	°C
D_t	Total drainage run-off in day t	mm
S_t	Total snow water equivalent in day t	mm
R_{t+1}	Precipitation registered day $t+1$ (mainly collected during day t)	mm
R_{3t}	Sum of precipitation last three days ($R_{t-2} + R_{t-1} + R_t$)	mm
S_Δ	Change in snow water equivalent ($S_t - S_{t-1}$)	mm

Table 1: Weather variables directly observed (upper part of the table) and derived (lower part).

Furthermore, we define $\boldsymbol{X}_k^\gamma$ as the reduced covariate matrix consisting only of the columns j of $\boldsymbol{X}_k$ for which $\gamma_{kj}^\lambda = 1$. The GLM for $\boldsymbol{\lambda}_k$ is given by

$$\begin{aligned}
\log(\boldsymbol{\lambda}_k) \mid \boldsymbol{\beta}_k, \sigma_k^2, \boldsymbol{\gamma}_{k\cdot}^\lambda, \boldsymbol{\zeta}_k &\sim \text{Normal}(\beta_{k0}\mathbf{1} + \boldsymbol{X}_k^\gamma \boldsymbol{\beta}_k^\gamma + \log(\boldsymbol{A}_k), \sigma_k^2 I) \\
\boldsymbol{\beta}_k(\boldsymbol{\gamma}_{k\cdot}^\lambda) \mid \sigma_k^2, \boldsymbol{\gamma}_{k\cdot}^\lambda.\boldsymbol{\zeta}_k &\sim \text{Normal}(\mathbf{0}, p_k\sigma_k^2({\boldsymbol{X}_k^\gamma}^T \boldsymbol{X}_k^\gamma)^{-1}) \\
p(\beta_{k0}) &\propto 1 \\
p(\sigma_k^2) &\propto \text{Inv-Gamma}(a, b),
\end{aligned} \tag{2}$$

where p_k is the number of days with $N_{kt} \geq 1$ and $\boldsymbol{A}_k$ is now the vector of insured units for those days only. The structure of the covariance of $\boldsymbol{\beta}_k^\gamma$ is in the form of a g-prior (Zellner, 1986), and has been considered for variable selection in Gaussian linear models (see e.g. Smith & Kohn, 1996; George & McCulloch, 1997; Fernandez *et al.*, 2001). In such models, the g-prior covariance structure replicates the covariance structure of the likelihood. The Gaussian-Inverse-Gamma conjugate prior structure is convenient, as it allows calculation of the marginal likelihood for the variable selection parameters, thus enabling sampling the variable selection indicator variables directly. This avoids the difficulty of the variable dimension of the parameter space. In our case, the Gaussian linear regression is lifted one level up in the hierarchy to model $\log(\lambda_{kt})$, which itself is a parameter in the Poisson Hurdle model. Thus the g-prior does not mimic the covariance structure of the likelihood, but has the same convenient properties. Our choice of p_k as the scale factor corresponds to unit prior information for $\boldsymbol{\beta}_k^\gamma$ as in Kass & Wasserman (1995), Kohn *et al.* (2001) and Smith & Fahrmeir (2007).

Integrating out the regression coefficients and the overdispersion variance results in the following prior for $\log(\boldsymbol{\lambda}_k)$

$$p(\log(\boldsymbol{\lambda}_k) \mid \boldsymbol{\gamma}_{k\cdot}^\lambda, \boldsymbol{\zeta}_k) \propto \left(1 + S_k(\boldsymbol{\theta}_k, \boldsymbol{\gamma}_{k\cdot}^\lambda)\right)^{-(p_k-1)/2} (1+p_k)^{-r_k/2} \tag{3}$$

where $\boldsymbol{\theta}_k = \log(\boldsymbol{\lambda}_k) - \log(\boldsymbol{A}_k)$, $r_k = \sum_{j=1}^q \gamma_{kj}^\lambda$ is the number of non-zero regression coefficients and

$$S_k(\boldsymbol{\theta}_k, \boldsymbol{\gamma}_{k\cdot}^\lambda) = (\boldsymbol{\theta}_k - \overline{\boldsymbol{\theta}}_k\mathbf{1})^T(I + p_k\boldsymbol{H}_k^\gamma)^{-1}(\boldsymbol{\theta}_k - \overline{\boldsymbol{\theta}}_k\mathbf{1}),$$

with $\overline{\boldsymbol{\theta}}_k$ being the average of the elements of $\boldsymbol{\theta}_k$ and

$$\boldsymbol{H}_k^\gamma = \boldsymbol{X}_k^\gamma({\boldsymbol{X}_k^\gamma}^T\boldsymbol{X}_k^\gamma)^{-1}{\boldsymbol{X}_k^\gamma}^T.$$

See Appendix A.1.1 for a full derivation of the model.

We make the variable selection for covariate j smooth across Norway by assuming apriori a spatial model for γ^λ. Define the $K \times q$ matrix of binary indicator random variables

$$\boldsymbol{\gamma}^\lambda = \begin{pmatrix} \gamma_{11}^\lambda & \cdot & \cdot & \cdot & \gamma_{1q}^\lambda \\ \vdots & \cdot & & & \vdots \\ \gamma_{k1}^\lambda & & \cdot & & \gamma_{kq}^\lambda \\ \vdots & & & \cdot & \vdots \\ \gamma_{k1}^\lambda & \cdot & \cdot & \cdot & \gamma_{Kq}^\lambda \end{pmatrix} = \begin{pmatrix} {\boldsymbol{\gamma}_{1\cdot}^\lambda}^T \\ \vdots \\ {\boldsymbol{\gamma}_{k\cdot}^\lambda}^T \\ \vdots \\ {\boldsymbol{\gamma}_{K\cdot}^\lambda}^T \end{pmatrix} = \left(\boldsymbol{\gamma}_{\cdot 1}^\lambda \cdots \boldsymbol{\gamma}_{\cdot j}^\lambda \cdots \boldsymbol{\gamma}_{\cdot q}^\lambda\right),$$

We assume a spatial model for each covariate j across Norway, by giving $\boldsymbol{\gamma}^{\lambda}_{\cdot j} = (\gamma^{\lambda}_{1j}, \ldots, \gamma^{\lambda}_{Kj})^T$ (the vector of indicators for covariate j over all municipalities K) an Ising prior distribution

$$p(\boldsymbol{\gamma}^{\lambda}_{\cdot j} \mid \omega_j) \propto \exp\left(\omega_j \sum_{k' \sim k} I(\gamma^{\lambda}_{k'j} = \gamma^{\lambda}_{kj})/d_{k'k}\right),\ j = 1, \ldots, q.$$

with ω_j apriori uniformly distributed on $(0, \omega_{\text{max}})$ for some fixed value of ω_{max}. Here $k' \sim k$ indicates that the two municipalities k and k' are neighbours, with a distance of $d_{k'k}$. Various topologies are possible. Here we simply assume that municipalities sharing a boundary are neighbours, and $d_{k'k} = 1, \forall (k', k)$. Smith & Fahrmeir (2007) used the Ising model to spatially smooth the variable selection process in linear regression models on a regular lattice. The q variable selection indicator variable vectors $\boldsymbol{\gamma}^{\lambda}_{\cdot j}$, $j = 1, \ldots, q$ are assumed apriori to be independent.

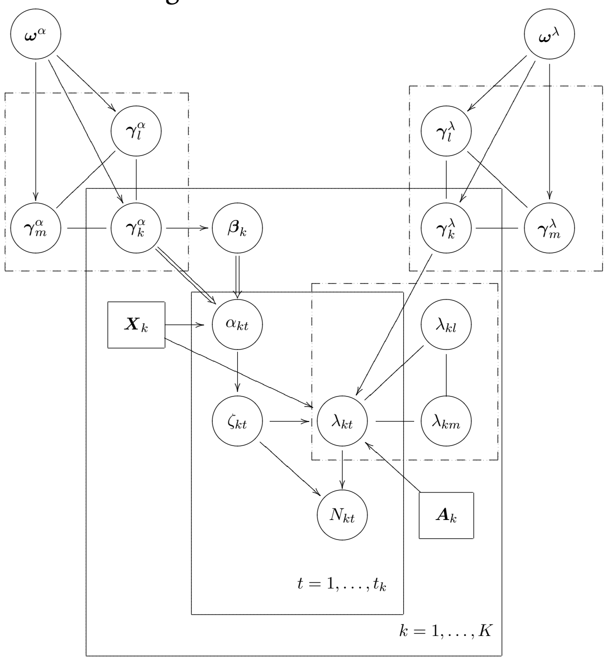
We move now to the other component of the model, describing presence or absence of claims. The variable selection in the GLM for α_{kt} is done in the same way as for $\boldsymbol{\lambda}_k$, using the variable selection indicator matrix $\boldsymbol{\gamma}^{\alpha}$. For each municipality k, $\gamma^{\alpha}_{kj} = 1$ means that covariate j enters the model for α_{kt}. Abusing notation, we now let $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kq})^T$ be the vector of regression coefficients for the GLM for α_{kt} and define $\boldsymbol{\beta}^{\gamma}_k$ as the reduced vector of β_{kj} for which $\gamma^{\alpha}_{kj} = 1$. Here, $\boldsymbol{X}_k$ is again the full covariate matrix for all the days $t \in T_k$. The GLM for $\boldsymbol{\alpha}_k$ is given by

$$\begin{aligned} \text{logit}(\boldsymbol{\alpha}_k) &= \beta_{k0}\mathbf{1} + \boldsymbol{X}^{\gamma}_k \boldsymbol{\beta}^{\gamma}_k \\ \boldsymbol{\beta}^{\gamma}_k \mid \boldsymbol{\gamma}^{\alpha}_{k\cdot} &\sim \text{Normal}(\mathbf{0}, 4t_k(\boldsymbol{X}^{\gamma T}_k \boldsymbol{X}^{\gamma}_k)^{-1}) \\ \beta_{k0} &\sim N(0, 4) \end{aligned} \tag{4}$$

where $\text{logit}(\boldsymbol{\alpha}_k)$ is the vector of components $\text{logit}(\alpha_{kt})$ for all t. Details on the choice of priors for the regression coefficients can be seen in Appendix A.1.2. The prior on $\boldsymbol{\gamma}^{\alpha}$ is exactly the same as the prior on $\boldsymbol{\gamma}^{\lambda}$, with different hyperparameters ω^{α}_j.

A graph representation of the complete model is given in Figure 2. Sampling from the posterior distributions can be done separately for the two model parts, see Appendix A.2 for details on the algorithms.

Figure 2: Graph representation of the full model. Square nodes represent data. The node α_{kt} is a deterministic function of its parents, which is indicated by the double arrows ($\Rightarrow$). Here l and m represent municipalities which are neighbours to k.



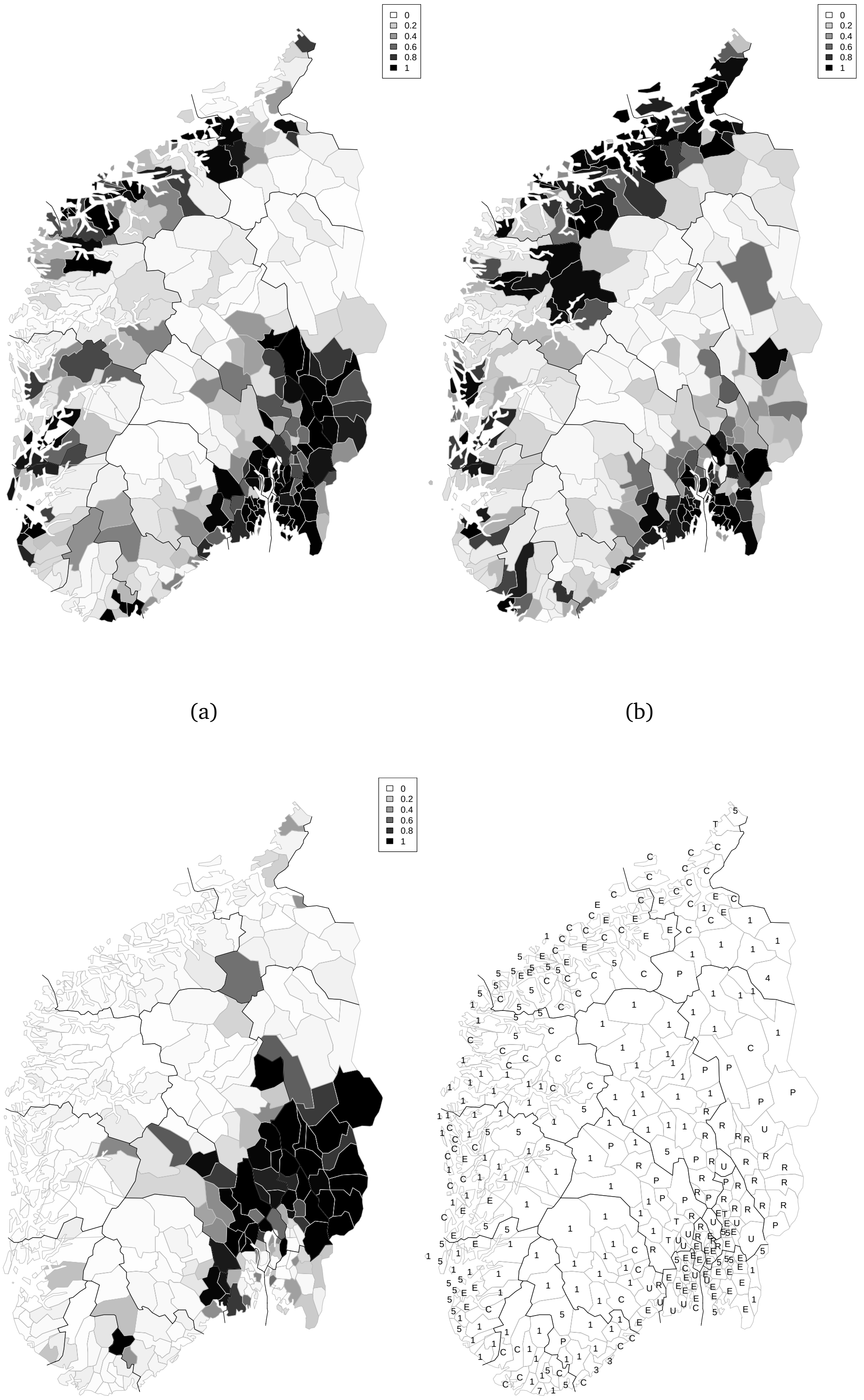
4 Results

To investigate the predictive abilities of our model and study the fit to the data, we divided the data into a training set, used for the posterior analysis, and a test set, reserved for evaluating predictions. The test set consists of one of the ten years of data (the year 2001), and the training set contains of the other 365×9 days.

What are the covariates that appear a posteriori significantly in the first component of the Hurdle model, concentrating on presence/absence of claims? Of the seven weather covariates (see Table 1), four appear to have no, or very little effect: Temperature (C_t), snow-water equivalent (S_t) and snow difference (S_Δ) and the mean precipitation during the last three days (R_{t3}) (results not shown). The other three covariates (drainage run-off D_t, precipitation on the previous day and early morning R_t, precipitation on the same day R_{t+1}) have an important role. Figure 3 a illustrates the effect of R_{t+1}, showing for each municipality k the Monte Carlo estimate of the posterior probability of $\gamma^{\alpha}_{kj} = 1$. Same-day precipitation is important for most of the western coast (though not around the Sognefjord) and in south-east Norway, around the Oslofjord, but less along the south coast, and not at all in the mountainous central areas, where exposure is very low. Also, the effect is present along the southern border to Sweden (south Hedmark). In general, precipitation has less impact in mountainous and remote areas than in urbanised areas due to vegetation and soil absorbing water, as opposed to asphalt covered streets in towns which rely on a properly dimensioned sewage system to collect the water. Figure 3 b illustrates the effect of the precipitation in the previous day R_t. Comparing these two maps, we see that the effect of the previous day is stronger and more widespread along the north-west coast, penetrating into the country, but still stops when the altitude begins to increase. The effect of drainage (Figure 3 c) is very strong in south-east Norway, just off the coast, and for most municipalities below the mountains. These areas are flatter and water does not escape as easily. The map in Figure 3 d indicates which model has largest posterior probability among the 128 possibilities (the coding of which covariates are in the different models can be found in Table 3 in Appendix A.3). Inspecting the posterior probabilities of the models for each municipality shows that the most likely model very often has a probability much larger than all other models (L-shaped density). Eleven models are present in this map of most probable models, but the most frequent ones involve some (or all)of R_{t+1}, R_t and D_t. On the west coast, precipitation alone enters the most probable model, while in south-east Norway, more covariates are suggested. For example, in Bergen where snow is rare and terrain varied, the selected model includes R_t and R_{t+1}.

Next we look to the results for the second component of the Hurdle model, the positive count part, and investigate the posterior probability of $\gamma^{\lambda}_{kj} = 1$. First we consider the a posteriori most probable model, and compare Figure 4 c and Figure 3 d. The models are much more rich in covariates, and differ more, including locally. Snow variables are often present, both snow equivalent on the same day and change in snow equivalent in the last two days (a measure of melting snow). For example, in Bergen the selected model includes R_t, R_{t+1}, S_t and S_Δ. Not all municipalities have counts > 1, which means that there is no Positive Poisson model for these municipalities and hence the spatial models for the γ^{λ}_{kj}'s are less smooth than was the case for the γ^{α}_{kj}'s. While precipitation and drainage are important for conditions that lead to at least one claim, the number of claims (that is the seriousness of the local damage) seems to be modulated by more complex weather events. Of course, this may to some extent be explained by the fact that there is less data for the λ_k model than for the α_k model. The variables which have large posterior probability of $\gamma^{\lambda}_{kj} = 1$ are precipitation on the same day and the day before: these maps are very similar to those seen for the γ^{α}_{kj}'s (results not shown). Drainage has no importance for the quantity of claims (results not shown), while it does have importance for the presence of claims, therefore it must be associated more to isolated weather related damages. Snow equivalent on the same day S_t (Figure 4 a) and the difference in snow equivalent S_Δ (Figure 4 b) enter now to a varying extent all over the country; S_t with highest probability mostly along the coast, and S_Δ mostly on the border between the regions Oppland and Hedmark. On the west coast, snow is quite rare. When it snows, the temperatures are mostly around freezing, and the snow is wet and heavy which can make water collection systems dense. Along the south-coast, snow often comes in extremely heavy amounts over short time intervals.

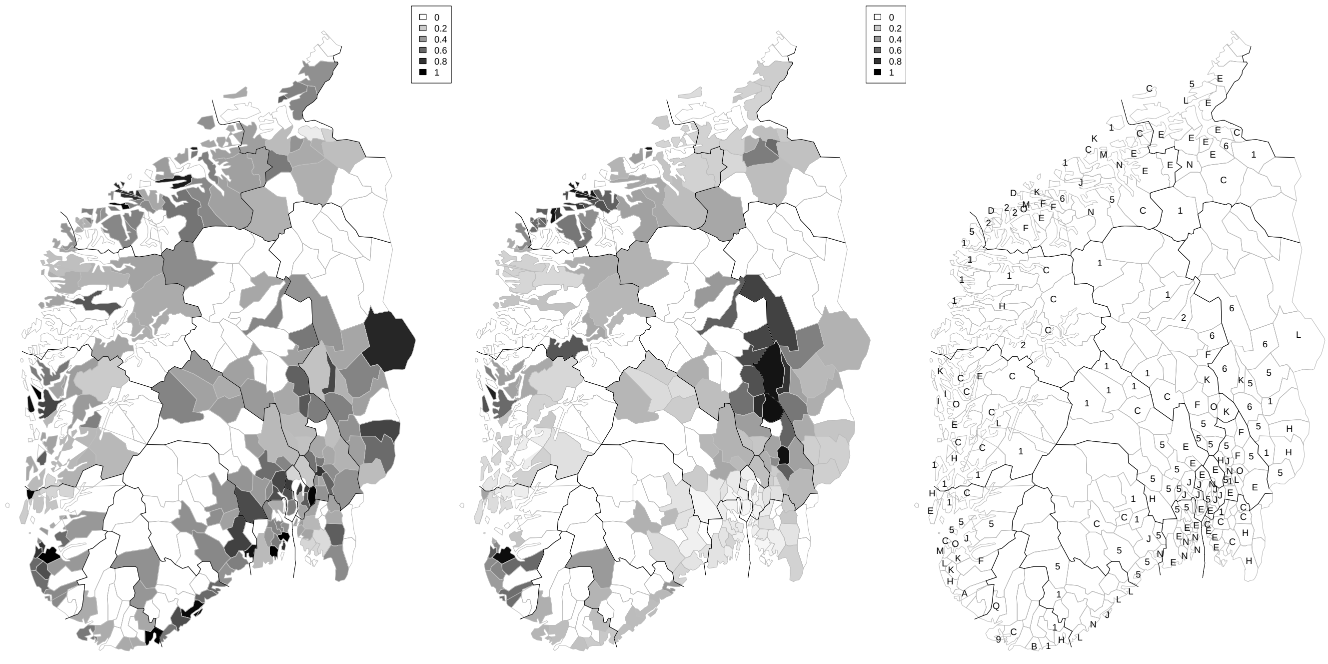
Figure 3: Maps of south and central Norway, divided into the municipalities, showing the Monte Carlo estimate of the posterior probability of the binary inclusion variable $\gamma_{kj}^{\alpha} = 1$ for each municipality k for covariate j representing (a) precipitation on the current day R_{t+1}, (b) precipitation on the previous day and early morning R_t and (c) drainage run-off D_t. A map of the model with largest posterior probability, among the 128 possible ones, for each municipality is shown in (d).



(a) (b)

(c) (d)

Figure 4: Maps of south and central Norway, divided into the municipalities, showing the Monte Carlo estimate of the posterior probability of the binary inclusion variable $\gamma_{kj}^{\lambda} = 1$ for each municipality k for covariate j representing (a) snow equivalent on the same day S_t and (b) difference in snow equivalent S_Δ. A map of the model with largest posterior probability, among the 128 possible ones, for each municipality is shown in (c). For 118 of the 319 municipalities almost all (and in many cases all) positive observed counts are equal to 1, and hence it is not possible to fit a Positive Poisson part. The model for these municipalities is collapsed to the binary model of a 0 or 1 count. Some municipalities have too few positive counts to fit the full GLM for λ_k, and hence only the intercept is included in the λ_k model for these municipalities. For 5 municipalities, the covariate $S_\Delta = S_t - S_{t-1}$ is linearly dependent on the covariate S_t, and S_t is therefore not included for these municipalities.



(a) (b) (c)

4.1 Prediction

We used all data except for the year 2001, to predict the number of claims for year 2001. The posterior predictive distribution and details on how to sample from it can be found in Appendix A.4. As weather predictions are considered reliable for one-week-ahead, we studied how well we could predict the number of claims in each municipality in each of the 52 weeks of 2001. We used actual observed weather, instead of weather predictions, to compute the posterior predictive distribution for the weekly number of claims for each municipality in 2001.

In order to evaluate the predictive performance, each week of 2001 is classified as one of three types: 'Week Type 0'=No claims, 'Week Type 1'=1, 2, or 3 claims and 'Week Type 2'=4 or more claims (nationwide, this is approximately 5% of the weeks). The type of each week of 2001 for each municipality was predicted as the type with the highest posterior predictive probability using the posterior predictive distribution of the number of claims. On average, the percentage of the 52 weeks in 2001 with predicted class equal to observed class for a municipality is 89%. The "success" percentages for the four largest cities are: Oslo and Bergen, 69% of weeks, Trondheim, 71% of weeks and Stavanger, 67% of weeks. With 46%, Sarpsborg is the only municipality with $< 50\%$ of weeks with predicted class equal to observed class.

To investigate how predictions are in extreme situations, we considered for each municipality the four weeks among all 52 weeks in 2001 with the highest observed number of claims, i.e. the four weeks with the maximum observed values of $\sum_{t \in \text{a week}} N_{kt}$. The posterior predictive quantiles of $\sum_{t \in \text{a week}} N_{kt}$, together with the observed number of claims, for those four weeks for Oslo and Bergen can be seen in Table 2. There is a tendency to underpredict the number of claims. For comparison, Table 2 also displays the corresponding prediction results for the four weeks with medial observed number of claims: predictions are excellent, with no sign of systematic bias. A different comparison is described in the bottom half of Table 2. Here we consider the four weeks among all weeks in 2001 with the maximum total precipitation, and also the four weeks with the medial total precipitation. Comparing the prediction results for the most rainy weeks with the prediction results for the medial rainy weeks in Bergen and Oslo, we see less evident negative bias than for the comparison between the prediction results for the weeks with the maximum and medial observed number of claims. Our model can apparently cope reasonably well with extreme precipitations, but is less able to predict extreme numbers of claims. One possible reason can be that we lack one or more weather covariates that cause extreme number of claims.

Table 2: Posterior predictive quantiles and actual observation of $\sum_{t \in \text{a week}} N_{kt}$ for (a) the four weeks with the maximum observed $\sum_{t \in \text{a week}} N_{kt}$, (b) the four weeks with the median observed $\sum_{t \in \text{a week}} N_{kt}$, (c) the four weeks with maximum total precipitation and (d) the four weeks with medial total precipitation, for Oslo and Bergen.

	Oslo						Bergen					
	Posterior predictive quantiles					Observed	Posterior predictive quantiles					Observed
	0%	25%	50%	75%	100%	$\sum N_{kt}$	0%	25%	50%	75%	100%	$\sum N_{kt}$
(a)	0	3	4	7	54	11	0	2	3	4	13	7
	0	3	4	6	18	11	0	2	3	4	12	7
	0	2	3	5	17	8	0	1	2	3	9	6
	0	2	3	4	14	7	0	1	2	4	11	6
(b)	0	2	3	4	34	3	0	1	2	4	10	2
	0	2	3	4	10	3	0	1	2	4	10	2
	0	2	3	4	10	3	0	2	3	4	17	2
	0	2	3	4	15	3	0	1	2	3	9	2
(c)	0	3	5	7	28	5	0	3	4	6	18	5
	0	3	4	7	54	11	0	3	4	6	17	1
	0	2	4	5	30	6	0	2	3	5	15	3
	0	3	4	6	18	11	0	2	3	5	15	3
(d)	0	2	3	4	12	6	0	2	3	4	11	0
	0	1	3	4	16	3	0	1	3	4	11	2
	0	2	3	4	16	1	0	1	2	4	10	3
	0	2	3	4	15	3	0	1	3	4	11	3

5 Discussion

In this paper we have developed a new statistical approach to explain and predict insurance losses based on weather events on a local scale. We have considered the number of claims; similar models can be derived for the type of damage and its economical value. In this case mixed gamma models are used (Yip & Yau, 2005), conditioned on damage happening ($N_{kt} > 0$). In our model, we separate the occurrence of claims in each municipality and day, from the actual number of claims therein. Results indicate that there are differences in which weather covariates explain and best predict these dynamics, however this difference may be partly due to the fact that there is less data available to fit the model for the positive counts. We have suggested a Bayesian spatially smooth variable selection approach, assuming local spatial homogeneity in the underlying meteorological causes. As is usual in spatial Bayesian inference, this strengthens inference and prediction, as areas with less data can borrow strength from neighbouring ones with more data. Incorporating geographical gradients to improve neighbourhood structure is possible. We smooth the latent variable selection variables γ^α and γ^λ, not the regression coefficients. We do not expect the regression coefficients to be smooth in space, as the strength of the effect of the covariates depends more on local factors.

Our study of damages to buildings due to externally inflicted water damage shows interesting regional patterns. Finding the weak points of regional building traditions allows petitioning for improvements, both with owners and local authorities. Mitigation and prevention are important strategies for the insurance industry. Our results can also be used to price policies better. We show that our model has sufficient predictive power to be useful in predicting high risk situations for damages to buildings based on short term weather forecasts. Also, it can predict with sufficient precision the regional distribution of damages immediately after a weather event, allowing a more efficient dispatchment of insurance inspectors. This study shows that weather information in Norway is useful in the near-the-event market.

In a broader perspective, our study can be combined with future climate predictions in order to estimate possible changes in the frequency and number of claims (Haug *et al.*, 2008, 2009). Downscaled regional climate models provide scenarios of temperature, precipitation, and many other weather variables at a fine scale (e.g. 10 by 10 km) for the decades to come, under various hypotheses. We can use these climate predictions as input in our model to predict, for example, the distribution of the yearly number of claims in every municipality in the future. The reliability of regional climate prediction is, however, unclear, especially in countries like Norway that have important mountain chains. These are poorly represented in the rough global circulation models which deliver the input for downscaling. We are currently working on evaluating downscaling techniques for Norway, comparing predicted and empirical distributions of precipitation in the past. When reliable downscaling at a fine scale is validated, we will use it to provide the insurance industry with predictions of exposure to climate risk. Most likely, as the underlying statistical distribution of events will change, this will lead to new adaptation strategies for the insurance industry.

A Appendix

A.1 Detailed model description

A.1.1 Details on the model for λ_{kt}

We define the vector $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kq})^T$ to be the coefficients for the covariates for the GLM for λ_{kt}. Let $\beta_{kj} = 0$ if $\gamma^{\lambda}_{kj} = 0$, $j = 1, \ldots, q$, while β_{kj} has a continuous distribution if $\gamma^{\lambda}_{kj} = 1$, and define $\boldsymbol{\beta}^{\gamma}_k$ as the vector of β_{kj} for which $\gamma^{\lambda}_{kj} = 1$. The intercept β_{k0} is always part of the model. Denote now by ${}_p\boldsymbol{X}_k$ the submatrix consisting of the rows of $\boldsymbol{X}_k$ corresponding to the days with positive count. All the columns of ${}_p\boldsymbol{X}_k$ (except the intercept column) are centred and scaled. The intercept is hence orthogonal to the covariates. Conditional on $\gamma^{\lambda}_{k\cdot}$ and $\boldsymbol{\zeta}_k$, the GLM for $\boldsymbol{\lambda}_k$ is given by

$$\begin{aligned} \log(\boldsymbol{\lambda}_k) \mid \boldsymbol{\beta}_k, \sigma^2_k, \boldsymbol{\gamma}^{\lambda}_{k\cdot}, \boldsymbol{\zeta}_k &\sim \text{Normal}(\beta_{k0}\mathbf{1} + {}_p\boldsymbol{X}^{\gamma}_k \boldsymbol{\beta}^{\gamma}_k + \log(\boldsymbol{A}_k), \sigma^2_k I) \\ \boldsymbol{\beta}^{\gamma}_k \mid \sigma^2_k, \boldsymbol{\gamma}^{\lambda}_{k\cdot}.\boldsymbol{\zeta}_k &\sim \text{Normal}(\boldsymbol{\mu}^{\gamma}_k, p_k \sigma^2_k ({}_p\boldsymbol{X}^{\gamma\,T}_k\, {}_p\boldsymbol{X}^{\gamma}_k)^{-1}) \\ p(\beta_{k0}) &\propto 1 \\ p(\sigma^2_k) &\propto \text{Inv-Gamma}(a, b), \end{aligned} \tag{5}$$

On the one hand the scale factor of the g-prior covariance should ensure an appropriately non-informative prior; on the other, it must not be too large, otherwise the null model with intercept only will tend to be selected. This is, as noted by Chipman *et al.* (2001), a form of the Bartlett-Lindley paradox (Bartlett, 1957). Our choice of p_k as the scale factor corresponds to unit prior information for $\boldsymbol{\beta}^{\gamma}_k$ (Kass & Wasserman, 1995), as used for variable selection in Gaussian linear models by e.g. Kohn *et al.* (2001) and Smith & Fahrmeir (2007). There are various possibilities for choosing the conditional mean $\boldsymbol{\mu}^{\gamma}_k$. We use here $\boldsymbol{\mu}^{\gamma}_k = \mathbf{0}$. The reason for giving β_{k0} a flat prior is that we avoid the prior guess of location of $\boldsymbol{\lambda}_k$ (Fernandez *et al.*, 2001; Bottolo & Richardson, 2008). Since the columns of ${}_p\boldsymbol{X}_k$ are centred and the intercept is orthogonal to the other covariates, this does not affect the conjugacy.

Because of the conjugate prior distributions for $\boldsymbol{\beta}^{\gamma}_k$ and σ^2_k in (5), we can obtain the density for $\log(\boldsymbol{\lambda}_k)$ conditioned only on $\gamma^{\lambda}_{k\cdot}$ and $\boldsymbol{\zeta}_k$ given in (3) in Section 3 by integrating out β_{k0}, $\boldsymbol{\beta}^{\gamma}_k$ and σ^2_k. This is analogous to deriving the marginal likelihood in the g-prior variable selection setting for Gaussian linear models (see e.g. Smith & Kohn (1996), George & McCulloch (1997), and Fernandez *et al.* (2001) for a flat prior on the intercept). To ease notation, we here drop the municipality index k, $\boldsymbol{\lambda}$ indication on $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, indication of dependency on $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{\lambda}_{k\cdot}$ and positive count indicator p. Hence $\boldsymbol{\lambda} = \boldsymbol{\lambda}_k$, $N = N_k$, $\boldsymbol{A} = \boldsymbol{A}_k$, $\boldsymbol{\zeta} = \boldsymbol{\zeta}_k$, $\boldsymbol{\beta} = \boldsymbol{\beta}^{\gamma}_k$, $\beta_0 = \beta_{k0}$, $\boldsymbol{\mu} = \boldsymbol{\mu}^{\gamma}_k$ and $\boldsymbol{X} = {}_p\boldsymbol{X}^{\gamma}_k$. First, well known conjugate Gaussian-Inverse-Gamma family calculus gives us

$$\begin{aligned} p(\beta_0, &\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\zeta}) = p(\log(\boldsymbol{\lambda}) \mid \beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\zeta}) p(\beta_0) p(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\zeta}) p(\sigma^2) \cdot A_1(\boldsymbol{\lambda}) \\ &= (\sigma^2)^{-p/2} \exp\left\{ -\tfrac{1}{\sigma^2} (\boldsymbol{\theta} - \beta_0\mathbf{1} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{\theta} - \beta_0\mathbf{1} - \boldsymbol{X}\boldsymbol{\beta}) \right\} \\ &\cdot \left|\sigma^2 \boldsymbol{\Sigma}^0\right|^{-1/2} \cdot \exp\left\{ -\tfrac{1}{\sigma^2} (\boldsymbol{\beta} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{0^{-1}} (\boldsymbol{\beta} - \boldsymbol{\mu}) \right\} \\ &\cdot \left(\sigma^2\right)^{-(a+1)} \cdot \exp\left\{ -b/\sigma^2 \right\} \cdot A_1(\boldsymbol{\lambda}) \cdot A_2 \\ &= \exp\Big\{ -\tfrac{1}{\sigma^2} \Big[\left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1}\right)^T \left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1}\right) + (1/p) \cdot \boldsymbol{\mu}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\mu} - \widehat{\boldsymbol{\beta}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\beta}} \\ &+ \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T \widehat{\boldsymbol{\Sigma}}^{-1} \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) + p\left(\beta_0 - \overline{\boldsymbol{\theta}}\right)^2 + 2b \Big] \Big\} \cdot \left|\sigma^2 p (\boldsymbol{X}^T \boldsymbol{X})^{-1}\right|^{-1/2} \\ &\cdot \left(\sigma^2\right)^{-(a+1+p/2)} \cdot A_1(\boldsymbol{\lambda}) \cdot A_2 \\ &= \exp\Big\{ -\tfrac{1}{\sigma^2} \Big[\left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1} - \boldsymbol{X}\boldsymbol{\mu}\right)^T (I - \boldsymbol{X}\widehat{\boldsymbol{\Sigma}}\boldsymbol{X}^T) \left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1} - \boldsymbol{X}\boldsymbol{\mu}\right) \\ &+ \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T \widehat{\boldsymbol{\Sigma}}^{-1} \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) + p\left(\beta_0 - \overline{\boldsymbol{\theta}}\right)^2 + 2b \Big] \Big\} \cdot \left|\sigma^2 p (\boldsymbol{X}^T \boldsymbol{X})^{-1}\right|^{-1/2} \\ &\cdot \left(\sigma^2\right)^{-(a+1+p/2)} \cdot A_1(\boldsymbol{\lambda}) \cdot A_2 \end{aligned} \tag{6}$$

where $\boldsymbol{\theta} = \log(\boldsymbol{\lambda}) - \log(\boldsymbol{A})$, $\overline{\boldsymbol{\theta}}$ is the mean of $\boldsymbol{\theta}$,

$$\begin{aligned}
\boldsymbol{\Sigma}^0 &= p(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\
\widehat{\boldsymbol{\Sigma}} &= \left(\boldsymbol{\Sigma}^{0^{-1}} + \boldsymbol{X}^T\boldsymbol{X}\right)^{-1} = (p/(p+1)) \cdot (\boldsymbol{X}^T\boldsymbol{X})^{-1} \\
\widehat{\boldsymbol{\beta}} &= \widehat{\boldsymbol{\Sigma}}\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\mu}/p + \boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1})
\end{aligned}$$

and

$$I - \boldsymbol{X}\widehat{\boldsymbol{\Sigma}}\boldsymbol{X}^T = I - \boldsymbol{X}\left(\boldsymbol{\Sigma}^{0^{-1}} + \boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T = (I + \boldsymbol{X}\boldsymbol{\Sigma}^0\boldsymbol{X}^T)^{-1} = (I + p\boldsymbol{H})^{-1},$$

with

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T.$$

The density of the p-variate t-distributed $\boldsymbol{y}$ with ν degrees of freedom, mean vector $\boldsymbol{\eta}$ and correlation matrix $\boldsymbol{R}$ (and hence covariance matrix $\boldsymbol{R} \cdot \nu/(\nu-2)$), denoted by $\text{MVT}_p(\nu, \boldsymbol{\eta}, \boldsymbol{R})$, is

$$f(\boldsymbol{y}) = \left[\nu + (\boldsymbol{y}-\boldsymbol{\eta})^T \boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{\eta})\right]^{-(\nu+p)/2} \cdot \frac{\nu^{\nu/2}\left(\frac{\nu+p}{2}-1\right)!}{\pi^{p/2}\left(\frac{\nu}{2}-1\right)!}\left|\boldsymbol{R}\right|^{-1/2} \tag{7}$$

From (6) and (7) we get

$$\begin{aligned}
p(\log(\boldsymbol{\lambda}) \mid \gamma, \boldsymbol{\zeta}) &= \int\!\!\int\!\!\int p(\log(\boldsymbol{\lambda}), \beta_0, \boldsymbol{\beta}, \sigma^2 \mid \gamma, \boldsymbol{\zeta}) \mathrm{d}\beta_0 \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \\
&= \int\!\!\int\!\!\int p(\log(\boldsymbol{\lambda}) \mid \beta_0, \boldsymbol{\beta}, \sigma^2, \gamma, \boldsymbol{\zeta}) p(\beta_0) p(\boldsymbol{\beta} \mid \sigma^2, \gamma, \boldsymbol{\zeta}) p(\sigma^2) \mathrm{d}\beta_0 \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \\
&\propto \int\!\!\int\!\!\int \exp\Big\{ -\tfrac{1}{\sigma^2}\Big[\left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1} - \boldsymbol{X}\boldsymbol{\mu}\right)^T (I - \boldsymbol{X}\widehat{\boldsymbol{\Sigma}}\boldsymbol{X}^T) \left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1} - \boldsymbol{X}\boldsymbol{\mu}\right) \\
&+ \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T \widehat{\boldsymbol{\Sigma}}^{-1} \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) + p\left(\beta_0 - \overline{\boldsymbol{\theta}}\right)^2 + 2b\Big]\Big\} \cdot \left|\sigma^2 p(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right|^{-1/2} \\
&\cdot \left(\sigma^2\right)^{-(a+1+p/2)} \mathrm{d}\beta_0 \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \\
&\propto \int \exp\Big\{ -\tfrac{1}{\sigma^2}\Big[\left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1} - \boldsymbol{X}\boldsymbol{\mu}\right)^T (I - \boldsymbol{X}\widehat{\boldsymbol{\Sigma}}\boldsymbol{X}^T) \left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1} - \boldsymbol{X}\boldsymbol{\mu}\right) + 2b\Big]\Big\} \\
&\cdot \left|\sigma^2 p(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right|^{-1/2} \cdot \left(\sigma^2\right)^{-(a+1+(p-1)/2)} \cdot \left|\sigma^2\widehat{\boldsymbol{\Sigma}}\right|^{1/2} \mathrm{d}\sigma^2 \\
&\propto \left(2b + S(\boldsymbol{\theta}, \boldsymbol{\gamma})\right)^{-(2a+p-1)/2} (1+p)^{-r/2} \\
&\propto \text{MVT}_p\left(2a, \log(\boldsymbol{A}) + \overline{\boldsymbol{\theta}}\mathbf{1} + \boldsymbol{X}\boldsymbol{\mu}, \tfrac{b}{a}(I + p\boldsymbol{H})\right),\ 2a > 0
\end{aligned} \tag{8}$$

where

$$S(\boldsymbol{\theta}, \boldsymbol{\gamma}) = (\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1} - \boldsymbol{X}\boldsymbol{\mu})^T (I + t\boldsymbol{H})^{-1} (\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1} - \boldsymbol{X}\boldsymbol{\mu})$$

and

$$r = \sum_{j=1}^{q} \boldsymbol{\gamma}_{kj}$$

(the number of non-zero regression coefficients). Notice that

$$I + p\boldsymbol{H} = (I - (p/(p+1))\boldsymbol{H})^{-1}.$$

When $r = 0$ and the model consists of intercept only, it is easy to see that integrating out β_0 and σ^2 gives

$$\begin{aligned}
p(\log(\boldsymbol{\lambda}) \mid \gamma, \boldsymbol{\zeta}) &\propto \left(2b + (\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1})^T(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\mathbf{1})\right)^{-(2a+p-1)/2} \\
&\propto \text{MVT}_p\left(2a, \log(\boldsymbol{A}) + \overline{\boldsymbol{\theta}}\mathbf{1}, \tfrac{b}{a}I\right),\ 2a > 0
\end{aligned}$$

For the prior distribution of σ_k^2 we chose to use $a \to 0$ and $b = 0.5$, which gives a relatively flat prior, while $b = 0.5$ ensures a more numerically stable $p(\log(\boldsymbol{\lambda}_k) \mid \boldsymbol{\gamma}_{k\cdot}^{\lambda}, \boldsymbol{\zeta}_k)$ than the choice of $b \to 0$.

A.1.2 The prior for the regression coefficients for the α_k model

Letting $\boldsymbol{X}_k$ be the full covariate matrix for all the days $t \in T_k$, centred and scaled and with an additional first intercept column of 1's, the prior for $\boldsymbol{\beta}_k^\gamma$ is

$$\boldsymbol{\beta}_k^\gamma \mid \boldsymbol{\gamma}_{k\cdot}^\alpha \sim \text{Normal}(\boldsymbol{\mu}_k^\gamma, \boldsymbol{\Sigma}_0^\gamma)$$

where

$$\boldsymbol{\Sigma}_0^\gamma = t_k(\boldsymbol{X}_k^{\gamma T} \boldsymbol{W}_k(\boldsymbol{\gamma}_{k\cdot}^\alpha, \boldsymbol{\mu}_k^\alpha) \boldsymbol{X}_k^\gamma)^{-1}.$$

Here, $\boldsymbol{W}_k(\boldsymbol{\gamma}_{k\cdot}^\alpha, \boldsymbol{\beta}_k)$ is the matrix of the usual GLM weights applied to α_{kt}, which are given by

$$\boldsymbol{W}_k^{-1}(\boldsymbol{\gamma}_{k\cdot}^\alpha, \boldsymbol{\beta}_k) = \text{diag}(b''(\nu_{kt}))g'(\alpha_{kt})^2)$$

with $\nu_{kt} = \text{logit}(\alpha_{kt})$, $b(\nu_{kt}) = \log(1 + \exp(\nu_{kt}))$ and $g(\alpha_{kt}) = \text{logit}(\alpha_{kt})$, so that

$$\boldsymbol{W}_k(\boldsymbol{\gamma}_{k\cdot}^\alpha, \boldsymbol{\beta}_k) = \frac{\exp(\boldsymbol{X}_k^\gamma \boldsymbol{\beta}_k^\gamma)}{\left(1 + \exp(\boldsymbol{X}_k^\gamma \boldsymbol{\beta}_k^\gamma)\right)^2},$$

following the notation in McCullagh & Nelder (1989). Analogously to the covariance specification for $\boldsymbol{\beta}_k^\gamma$, the prior covariance structure and scale factor t_k provide unit prior information for $\boldsymbol{\beta}_k^\gamma$, as suggested by Ntzoufras *et al.* (2003) and Nott & Leonte (2004) for variable selection in GLMs. We use here $\boldsymbol{\mu}_k^\gamma = \boldsymbol{0}$, which means that $\boldsymbol{W}_k(\boldsymbol{\gamma}_{k\cdot}^\alpha, \boldsymbol{\mu}_k^\alpha) = 1/4$ and hence $\boldsymbol{\Sigma}_0^\gamma = 4t_k(\boldsymbol{X}_k^{\gamma T} \boldsymbol{X}_k^\gamma)^{-1}$. Because the columns of $\boldsymbol{X}_k$ (except the intercept column) are centred, the intercept is orthogonal to the covariates. The Normal prior on the intercept is justifiable in this situation because t_k is large, while in the $\boldsymbol{\lambda}_k$ model the scale factor p_k could be quite small.

A.2 Sampling from the posterior distributions

For the model of the positive counts we suggest a Gibbs sampling approach. Let T_k^p be the subset of days with positive counts in municipality k, with cardinality p_k. The relevant full conditional distributions are

$$\begin{aligned}
p(\log(\boldsymbol{\lambda}_k) \mid \cdot) &\propto p(\boldsymbol{N}_k \mid \boldsymbol{\alpha}_k, \boldsymbol{\lambda}_k) \cdot p(\log(\boldsymbol{\lambda}_k) \mid \boldsymbol{\gamma}_{k\cdot}^\lambda, \boldsymbol{\zeta}_k) \propto \qquad (9)\\
&\prod_{t \in T_k^p} \left[\alpha_{kt} \mathbb{1}_{N_{kt}=0} + (1-\alpha_{kt}) \frac{\lambda_{kt}^{N_{kt}}}{(\exp(\lambda_{kt}) - 1)} \mathbb{1}_{N_{kt} \geq 1}\right]\\
&\cdot \left(2b + S_k(\boldsymbol{\theta}_k, \boldsymbol{\gamma}_{k\cdot}^\lambda)\right)^{-(2a+p_k-1)/2} (1+p_k)^{-r_k/2}\\
&\propto \prod_{t \in T_k^p} \left[\frac{\lambda_{kt}^{N_{kt}}}{(\exp(\lambda_{kt}) - 1)} 1_{N_{kt} \geq 1}\right] \cdot \left(2b + S_k(\boldsymbol{\theta}_k, \boldsymbol{\gamma}_{k\cdot}^\lambda)\right)^{-(2a+p_k-1)/2}
\end{aligned}$$

$$\begin{aligned}
p(\boldsymbol{\gamma}^\lambda \mid \cdot) &\propto \prod_{k=1}^K p(\log(\boldsymbol{\lambda}_k) \mid \boldsymbol{\gamma}_k^\lambda) \prod_{j=1}^q p(\boldsymbol{\gamma}_{\cdot j}^\lambda \mid \omega_j) \qquad (10)\\
&\propto \exp\left\{\sum_{k=1}^K l_k(\boldsymbol{\gamma}_{k\cdot}^\lambda) + \sum_{j=1}^q \omega_j \sum_{k' \sim k} I(\gamma_{k'j}^\lambda = \gamma_{kj}^\lambda)\right\}\\
&\text{with } l_k(\boldsymbol{\gamma}_{k\cdot}^\lambda) = \log\left(p(\log(\boldsymbol{\lambda}_k) \mid \beta_{k0}, \boldsymbol{\gamma}_{k\cdot}^\lambda, \boldsymbol{\zeta}_k)\right)\\
&= -((2a + p_k - 1)/2) \log\left(2b + S_k(\boldsymbol{\theta}_k, \boldsymbol{\gamma}_{k\cdot}^\lambda)\right) - (r_k/2)\log(1+p_k) + \text{constant}
\end{aligned}$$

$$\begin{aligned}
p(\omega_j \mid \cdot) &\propto p(\boldsymbol{\gamma}_{\cdot j}^\lambda \mid \omega_j) \cdot I(0 < \omega_j < \omega_{\text{max}}) \qquad (11)\\
&\propto \frac{1}{B_j(\omega_j)} \exp\left(\omega_j \sum_{k' \sim k} I(\gamma_{k'j}^\lambda = \gamma_{kj}^\lambda)\right) \cdot I(0 < \omega_j < \omega_{\text{max}}),\ j = 1, \ldots q,
\end{aligned}$$

Sampling from the full conditional distribution of $\boldsymbol{\gamma}^\lambda$ is facilitated by the fact that the full conditional probability $p(\gamma_{kj}^\lambda = 1 \mid \cdot)$ is available in closed form. Let $\boldsymbol{\gamma}^\lambda(\gamma_{kj}^\lambda = i)$ denote $\boldsymbol{\gamma}^\lambda$ with the

(k, j)'th element equal to i, and similarly $\boldsymbol{\gamma}_k^\lambda(\gamma_{kj}^\lambda = i)$ denote $\boldsymbol{\gamma}_k^\lambda$ with the j'th element equal to i, where $i \in \{0, 1\}$. Because

$$p(\gamma_{kj}^\lambda \mid \boldsymbol{\gamma}_{-(k,j)}^\lambda, \boldsymbol{\omega}^\lambda, \boldsymbol{\lambda}) \propto p(\boldsymbol{\gamma}^\lambda \mid \cdot),$$

we have

$$p(\gamma_{kj}^\lambda = 1 \mid \cdot) = \frac{p(\boldsymbol{\gamma}^\lambda(\gamma_{kj}^\lambda = 1) \mid \cdot)}{p(\boldsymbol{\gamma}^\lambda(\gamma_{kj}^\lambda = 0) \mid \cdot) + p(\boldsymbol{\gamma}^\lambda(\gamma_{kj}^\lambda = 1) \mid \cdot)} = \frac{1}{1 + \frac{p(\boldsymbol{\gamma}^\lambda(\gamma_{kj}^\lambda = 0) \mid \cdot)}{p(\boldsymbol{\gamma}^\lambda(\gamma_{kj}^\lambda = 1) \mid \cdot)}} = \frac{1}{1 + u_{kj}},$$

where

$$u_{kj} = \exp\left\{ l_k(\boldsymbol{\gamma}_k^\lambda(\gamma_{kj}^\lambda = 0)) - l_k(\boldsymbol{\gamma}_k^\lambda(\gamma_{kj}^\lambda = 1)) + \omega_j \sum_{i \in \delta k} \left(1 - 2\gamma_{ij}^\lambda\right) \right\}$$

with $\delta k = \{i \mid i \sim k\}$ (Smith & Kohn, 1996; Smith & Fahrmeir, 2007).

The normalising constant

$$B_j\left(\omega_j\right) = \sum_{\gamma_{\cdot j}^\lambda} \exp\left(\omega_j \sum_{k' \sim k} I(\gamma_{k'j}^\lambda = \gamma_{kj}^\lambda) \right),$$

summing over all possible values of the vector $\boldsymbol{\gamma}_{\cdot j}^\lambda$, is needed in order to obtain posterior samples of ω_j, but is computationally difficult to obtain exactly for regions with many municipalities. Following Smith & Fahrmeir (2007) we use an off-line computed approximation based on the thermodynamic integration approach of Green & Richardson (2002). For $\omega_j > 0$, we have $B_j(\omega_j) \propto \exp\left(c_j(\omega_j)\right)$, where

$$c_j(\omega_j) = \int_0^{\omega_j} E(U(\boldsymbol{\gamma}_j^\lambda) \mid {\omega_j}')\mathrm{d}{\omega_j}', \quad U(\boldsymbol{\gamma}_j^\lambda) = \sum_{k' \sim k} I(\boldsymbol{\gamma}_{k'j}^\lambda = \boldsymbol{\gamma}_{kj}^\lambda). \tag{12}$$

We get an approximation to $c_j(\omega_j)$ by approximating the integral in (12). The algorithm can be described as follows

1. Choose a discrete grid ${\omega_j}^i, i = 1, \ldots, I$ on the interval $[0, \omega_{\max}]$
2. For each ${\omega_j}^i, i = 1, \ldots, I$
   (a) Simulate ${\gamma_{\cdot j}^\lambda}^{[l]}$ from $p(\gamma_{\cdot j}^\lambda \mid \omega_j = {\omega_j}^i)$ for $l = 1, \ldots, L$. This can be done by single-site Gibbs sampling. Let $\boldsymbol{\gamma}_{-(k,j)}^\lambda$ be $\gamma_{\cdot j}^\lambda$ with γ_{kj}^λ omitted. Sample ${\gamma_{kj}^\lambda}^{[l]}$ from the conditional distribution $p(\boldsymbol{\gamma}_{kj}^\lambda \mid \omega_j = {\omega_j}^i, \boldsymbol{\gamma}_{-(k,j)}^\lambda)$, given by

   $$p(\boldsymbol{\gamma}_{kj}^\lambda = 1 \mid \omega_j = {\omega_j}^i, \boldsymbol{\gamma}_{-(k,j)}^\lambda) = 1/(1 + v_{kj})$$

   where

   $$v_{kj} = \exp\left\{ {\omega_j}^i \sum_{i \in \delta k} \left(1 - 2\boldsymbol{\gamma}_{ij}^\lambda\right) \right\}, \ \delta k = \{i \mid i \sim k\}$$

   (see (13)).
   (b) Calculate

   $$E\left[U(\boldsymbol{\gamma}_{\cdot j}^\lambda) \mid \omega_j = {\omega_j}^i\right] \approx \frac{1}{L} \sum_{l=1}^{L} U\left({\boldsymbol{\gamma}_{\cdot j}^\lambda}^{[l]}\right)$$

3. Approximate the continuous function $E(U(\boldsymbol{\gamma}_j^\lambda) \mid {\omega_j}')$ over the interval $[0, \omega_{\max}]$ by B-spline-based numerical interpolation, and store the spline coefficients
4. $c_j(\omega_j)$ can now be approximated for any $\omega_j \in [0, \omega_{\max}]$ by evaluating the corresponding numerical integral of the B-spline interpolation from 3. above

For the model dedicated to presence or absence of events, we need to design a reversible jump sampling scheme (Green, 1995), because $\boldsymbol{\beta}_k^{\gamma}$ has a variable dimension depending on $\boldsymbol{\gamma}_{k\cdot}^{\alpha}$. In our algorithm, we propose and accept/reject $\boldsymbol{\gamma}_{k\cdot}^{\alpha}$ and $\boldsymbol{\beta}_k^{\gamma}$ jointly: first, a new $\boldsymbol{\gamma}_{k\cdot}^{\alpha}{}'$ is proposed (which can be a "null" move, i.e. $\boldsymbol{\gamma}_{k\cdot}^{\alpha}{}' = \boldsymbol{\gamma}_{k\cdot}^{\alpha}$); then given $\boldsymbol{\gamma}_{k\cdot}^{\alpha}{}'$, $\boldsymbol{\beta}_k^{\alpha}{}'(\boldsymbol{\gamma}_{k\cdot}^{\alpha}{}')$ is proposed. We use the same proposal distribution for $\boldsymbol{\beta}_k^{\alpha}{}'(\boldsymbol{\gamma}_{k\cdot}^{\alpha}{}')$ as in Nott & Leonte (2004). This was previously suggested by Gamerman (1997) for the MCMC simulation of the regression coefficients in a GLM without variable selection. It is essentially the result of a single step of an iteratively weighted least squares algorithm for finding the posterior mode and approximate posterior covariance matrix for the regression coefficients in a GLM, see West (1985). To simplify notation, we drop the α index on $\boldsymbol{\gamma}^{\alpha}$ in the remainder of this section.

The proposal distribution $q(\boldsymbol{\beta}_k^{\gamma}, \boldsymbol{\beta}_k^{\gamma''})$ for proposing $\boldsymbol{\beta}_k^{\gamma''}$ (a vector of proposals β'_{kj} for the covariates for which $\gamma'_{kj} = 1$) from $\boldsymbol{\beta}_k^{\gamma}$ and $\boldsymbol{\gamma}'_{k\cdot}$ is

$$q(\boldsymbol{\beta}_k^{\gamma}, \boldsymbol{\beta}_k^{\gamma''}) = \text{Normal}(m(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_{k\cdot}, \boldsymbol{\gamma}'_{k\cdot}), \boldsymbol{M}(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_{k\cdot}, \boldsymbol{\gamma}'_{k\cdot}))$$

with

where

$$\begin{aligned}
\boldsymbol{M}(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_{k\cdot}, \boldsymbol{\gamma}'_{k\cdot}) &= \left(\boldsymbol{\Sigma}_0^{\gamma'^{-1}} + \boldsymbol{X}_k^{\gamma'^T} \boldsymbol{W}_k(\boldsymbol{\gamma}_{k\cdot}, \boldsymbol{\beta}_k) \boldsymbol{X}_k^{\gamma'}\right)^{-1} \\
m(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_{k\cdot}, \boldsymbol{\gamma}'_{k\cdot}) &= \boldsymbol{M}\left(\boldsymbol{\Sigma}_0^{\gamma'^{-1}} \boldsymbol{\mu}_k^{\gamma'} + \boldsymbol{X}_k^{\gamma'^T} \boldsymbol{W}_k(\boldsymbol{\gamma}_{k\cdot}, \boldsymbol{\beta}_k) \rho_{kt}(\boldsymbol{\beta}_k^{\gamma})\right) \\
&= \boldsymbol{M} \boldsymbol{X}_k^{\gamma'^T} \boldsymbol{W}_k(\boldsymbol{\gamma}_{k\cdot}, \boldsymbol{\beta}_k) \rho_{kt}(\boldsymbol{\beta}_k^{\gamma})
\end{aligned}$$

where

$$\rho_{kt}(\boldsymbol{\beta}_k^{\gamma}) = \boldsymbol{X}_k^{\gamma} \boldsymbol{\beta}_k^{\gamma} + (1 - \zeta_{kt} - \alpha_{kt}(\boldsymbol{\beta}_k^{\gamma})) g'(\alpha_{kt}(\boldsymbol{\beta}_k^{\gamma}))$$

with

$$\alpha_{kt}(\boldsymbol{\beta}_k^{\gamma}) = g^{-1}(\boldsymbol{X}_k^{\gamma} \boldsymbol{\beta}_k^{\gamma}) = \text{logit}^{-1}(\boldsymbol{X}_k^{\gamma} \boldsymbol{\beta}_k^{\gamma}) = 1/(1 + \exp(-\boldsymbol{X}_k^{\gamma} \boldsymbol{\beta}_k^{\gamma})).$$

Because we accept or reject $\boldsymbol{\gamma}_{k\cdot}$ and $\boldsymbol{\beta}_k^{\gamma}$ with a joint acceptance probability, it does not make sense to propose γ_{kj} from its full conditional distribution as we did for γ_{kj}^{λ}. Instead, we propose $\boldsymbol{\gamma}_{k\cdot}$ by sampling from the conditional prior for γ_{kj} given $\boldsymbol{\gamma}_{-(k,j)}$. Because $p(\gamma_{kj} \mid \boldsymbol{\gamma}_{-(k,j)}, \boldsymbol{\omega}^{\alpha}) \propto p(\boldsymbol{\gamma} \mid \boldsymbol{\omega}^{\alpha})$, we have

$$p(\gamma_{kj} = 1 \mid \boldsymbol{\omega}^{\alpha}, \boldsymbol{\gamma}_{-(k,j)}) = \frac{p(\boldsymbol{\gamma}(\gamma_{kj} = 1) \mid \boldsymbol{\omega}^{\alpha})}{p(\boldsymbol{\gamma}(\gamma_{kj} = 0) \mid \boldsymbol{\omega}^{\alpha}) + p(\boldsymbol{\gamma}(\gamma_{kj} = 1) \mid \boldsymbol{\omega}^{\alpha})} = \frac{1}{1 + \frac{p(\boldsymbol{\gamma}(\gamma_{kj}=0) \mid \boldsymbol{\omega}^{\alpha})}{p(\boldsymbol{\gamma}(\gamma_{kj}=1) \mid \boldsymbol{\omega}^{\alpha})}} = \frac{1}{1 + v_{kj}},$$

where

$$v_{kj} = \exp\left\{\omega_j^{\alpha} \sum_{i \in \delta k} (1 - 2\gamma_{ij})\right\}. \tag{13}$$

(Kohn *et al.*, 2001; Smith & Fahrmeir, 2007). Proposing from this conditional prior means that the proposal terms and prior terms for $\boldsymbol{\gamma}_{k\cdot}$ cancel out in the joint acceptance probability for $\boldsymbol{\gamma}_{k\cdot}$ and $\boldsymbol{\beta}_k^{\gamma}$. Also, since all elements of $\boldsymbol{\beta}_k^{\gamma}$ are proposed, that is $u = \boldsymbol{\beta}_k{}'(\boldsymbol{\gamma}_{k\cdot})$ and $(\boldsymbol{\beta}_k{}'(\boldsymbol{\gamma}_{k\cdot}), u') = (u, \boldsymbol{\beta}_k^{\gamma})$, the Jacobian in the reversible jump acceptance rate equals 1. Hence, the joint acceptance probability for $\boldsymbol{\gamma}_{k\cdot}$ and $\boldsymbol{\beta}_k^{\gamma}$ becomes

$$\min\left\{1, \frac{\prod_{t:N_{kt}=0} \alpha_{kt}(\boldsymbol{\beta}_k^{\gamma''}) \cdot \prod_{t:N_{kt}>0}\left(1 - \alpha_{kt}(\boldsymbol{\beta}_k^{\gamma''})\right) \cdot p(\boldsymbol{\beta}_k^{\gamma''} \mid \boldsymbol{\gamma}'_{k\cdot}) \cdot q(\boldsymbol{\beta}_k^{\gamma''}, \boldsymbol{\beta}_k^{\gamma})}{\prod_{t:N_{kt}=0} \alpha_{kt}(\boldsymbol{\beta}_k^{\gamma}) \cdot \prod_{t:N_{kt}>0} (1 - \alpha_{kt}(\boldsymbol{\beta}_k^{\gamma})) \cdot p(\boldsymbol{\beta}_k^{\gamma} \mid \boldsymbol{\gamma}_{k\cdot}) \cdot q(\boldsymbol{\beta}_k^{\gamma}, \boldsymbol{\beta}_k^{\gamma''})}\right\}.$$

For a "null" move, i.e. $\boldsymbol{\gamma}_{k\cdot}{}' = \boldsymbol{\gamma}_{k\cdot}$, this equals an ordinary Metropolis-Hastings move for $\boldsymbol{\beta}_k^{\gamma}$.

Posterior sampling for the interaction parameter ω_j^{α} can be done by Gibbs sampling in exactly the same way as for ω_j. The full conditional distribution is the same, when we replace ω_j and $\boldsymbol{\gamma}_{\cdot j}^{\lambda}$ with ω_j^{α} and $\boldsymbol{\gamma}_{\cdot j}$.

A.3 Coding of which covariates are in the different models

Model	R_t	R_{t+1}	T_t	D_t	S_t	R_{t3}	S_Δ
1							
2							x
3						x	
4						x	x
5		x					
6		x					x
7		x				x	
8		x				x	x
9			x				
A			x				x
B		x	x				
C	x						
D	x						x
E	x	x					
F	x	x					x
G	x	x				x	
H					x		
I					x		x
J		x			x		
K		x			x		x
L	x				x		
M	x				x		x
N	x	x			x		
O	x	x			x		x
P				x			
Q				x		x	
R		x		x			
S		x		x		x	
T	x			x			
U	x	x		x			

Table 3: Coding of which covariates are in the models that appear in the maps in Figure 3 d and Figure 4 e. The intercept is in all the models (Model 1 consists of intercept only).

A.4 Posterior predictive distribution

As described in Section 3, because the zero count part and the positive count part of the model are conditionally independent given the data N_{kt}, we can treat the two model parts separately when performing the posterior sampling. However, not conditioning on N_{kt} the two model parts are not independent, and hence for predictive sampling of random N_{kt}, the two model parts cannot be treated separately.

Let $\widetilde{T}_k$ be the set of days t for which we wish to predict N_{kt}, $\widetilde{\boldsymbol{N}}_k$, $\widetilde{\boldsymbol{\zeta}}_k$ and $\widetilde{\boldsymbol{\alpha}}_k$ the vectors of N_{kt}, ζ_{kt} and α_{kt}, $t \in \widetilde{T}_k$. Denote by $\widetilde{T}_k^p$ the subset of days in $\widetilde{T}_k$ with positive counts in municipality k, i.e. $\widetilde{T}_k^p = \{t \in \widetilde{T}_k : \zeta_{kt} = 1\}$, and let $\widetilde{p}_k$ be the number of days in $\widetilde{T}_k^p$, i.e. $\widetilde{p}_k = \sum_{t \in \widetilde{T}_k} \zeta_{kt}$. Now, $\log(\widetilde{\boldsymbol{\lambda}}_k)$ and $\log(\widetilde{\boldsymbol{A}}_k)$ are the vectors of $\log(\lambda_{kt})$ and $\log(A_{kt})$, $t \in \widetilde{T}_k^p$, $\widetilde{\boldsymbol{X}}_k$ is the covariate matrix for municipality k for $t \in \widetilde{T}_k$, and ${}_p\widetilde{\boldsymbol{X}}_k$ is the submatrix of the test data covariate matrix corresponding to the days $t \in \widetilde{T}_k^p$. Also, denote here by ${}_p\boldsymbol{X}_k$ the submatrix of the training data covariate matrix corresponding to the days with positive count $t \in T_k^p$. The columns of $\widetilde{\boldsymbol{X}}_k$ (except the intercept column) are "centred and scaled" by using the same centre and scaling factors used to produce the training covariate matrix $\boldsymbol{X}_k$ (which was used for the posterior analysis). Similarly, the columns of ${}_p\widetilde{\boldsymbol{X}}_k$ are "centred and scaled" by using the same centre and scaling factors used to produce ${}_p\boldsymbol{X}_k$.

To ease notation, we once again drop the municipality index k. The posterior predictive distribution of $\widetilde{\boldsymbol{N}}_k$ given the observed data $\boldsymbol{N} = \boldsymbol{N}_k$ is given by the following equations

$$
\begin{aligned}
p(\widetilde{\boldsymbol{N}} \mid \boldsymbol{N}) &= \int\int p(\widetilde{\boldsymbol{N}} \mid \widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\zeta}}) p(\widetilde{\boldsymbol{\zeta}} \mid \boldsymbol{N}) p(\log(\widetilde{\boldsymbol{\lambda}}) \mid \boldsymbol{N}, \widetilde{\boldsymbol{\zeta}}) \mathrm{d}\widetilde{\boldsymbol{\zeta}} \mathrm{d}\log(\widetilde{\boldsymbol{\lambda}}) \\
p(\widetilde{\boldsymbol{N}} \mid \widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\zeta}}) &= \prod_{t \in \widetilde{T}_k} p(N_{kt} \mid \lambda_{kt}, \zeta_{kt}) \\
p(N_{kt} = n \mid \lambda_{kt}, \zeta_{kt}) &= \begin{cases} \delta(0) & \text{if } \zeta_{kt} = 0 \\ \frac{\lambda_{kt}^n}{(\exp(\lambda_{kt}) - 1)n!} & \text{if } \zeta_{kt} = 1 \end{cases} \\
p(\widetilde{\boldsymbol{\zeta}} \mid \boldsymbol{N}) &= \sum_{\boldsymbol{\gamma}^\alpha} \int p(\widetilde{\boldsymbol{\zeta}} \mid \boldsymbol{\beta}^\gamma, \boldsymbol{\gamma}^\alpha) p(\boldsymbol{\beta}^\gamma, \boldsymbol{\gamma}^\alpha \mid \boldsymbol{N}) \mathrm{d}\boldsymbol{\beta}^\gamma \\
p(\widetilde{\boldsymbol{\zeta}} \mid \boldsymbol{\beta}^\gamma, \boldsymbol{\gamma}^\alpha) &= \text{Bernoulli}(1 - \widetilde{\boldsymbol{\alpha}}), \quad \widetilde{\boldsymbol{\alpha}} = \widetilde{\boldsymbol{X}}^\gamma \boldsymbol{\beta}^\gamma \\
p(\log(\widetilde{\boldsymbol{\lambda}}) \mid \boldsymbol{N}, \widetilde{\boldsymbol{\zeta}}) &= \sum_{\boldsymbol{\gamma}^\lambda} \int p(\log(\widetilde{\boldsymbol{\lambda}}) \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}^\lambda, \widetilde{\boldsymbol{\zeta}}) p(\log(\boldsymbol{\lambda}), \boldsymbol{\gamma}^\lambda \mid \boldsymbol{N}) \mathrm{d}\log(\boldsymbol{\lambda}) \\
p(\log(\widetilde{\boldsymbol{\lambda}}) \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}^\lambda, \widetilde{\boldsymbol{\zeta}}) &= \text{MVT}_{\widetilde{p}}\big(2a + p - 1, \log(\widetilde{\boldsymbol{A}}) + \overline{\boldsymbol{\theta}}\mathbf{1} + {}_p\widetilde{\boldsymbol{X}}^\gamma {}_\lambda\widehat{\boldsymbol{\beta}}^\gamma, \\
&\qquad \tfrac{2b + S(\theta, \gamma^\lambda)}{2a + p - 1}\big(I + \tfrac{1}{p}\mathbf{1}\mathbf{1}^T + \tfrac{p}{p+1}{}_p\widetilde{\boldsymbol{X}}^\gamma ({}_p\boldsymbol{X}^{\gamma T} {}_p\boldsymbol{X}^\gamma)^{-1} {}_p\widetilde{\boldsymbol{X}}^{\gamma T}\big)\big)
\end{aligned} \tag{14}
$$

where

$$
{}_\lambda\widehat{\boldsymbol{\beta}}^\gamma = (p/(p+1))({}_p\boldsymbol{X}^{\gamma T} {}_p\boldsymbol{X}^\gamma)^{-1} {}_p\boldsymbol{X}^{\gamma T}(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}).
$$

The full conditional (14) is obtained in the same way as in (8) by replacing $p(\log(\boldsymbol{\lambda}), \beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\gamma}, \boldsymbol{\zeta})$ with $p(\log(\widetilde{\boldsymbol{\lambda}}), \beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\zeta}})$. To ease notation further, we drop in the following the λ indicator on $\boldsymbol{\gamma}^\lambda$, the positive count indicator p on ${}_p\boldsymbol{X}$, ${}_p\widetilde{\boldsymbol{X}}$, as well as the γ indicator on $\boldsymbol{\beta}^\gamma$, $\boldsymbol{X}^\gamma$ and $\widetilde{\boldsymbol{X}}^\gamma$. From (6) it is easily seen that

$$
\begin{aligned}
p(\beta_0 \mid \boldsymbol{\lambda}, \sigma^2, \widetilde{\boldsymbol{\zeta}}) &\propto p(\beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\zeta}}) \propto \text{Normal}(\overline{\boldsymbol{\theta}}, \sigma^2/p) \\
p(\boldsymbol{\beta} \mid \boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\zeta}}) &\propto p(\beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\zeta}}) \propto \text{Normal}(\widehat{\boldsymbol{\beta}}, \sigma^2\widehat{\boldsymbol{\Sigma}}) \\
p(\sigma^2 \mid \boldsymbol{\lambda}, \widetilde{\boldsymbol{\zeta}}) &= \int\int p(\beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\zeta}}) \mathrm{d}\beta_0 \mathrm{d}\boldsymbol{\beta} = \text{Inv-Gamma}(a_p, b_p)
\end{aligned} \tag{15}
$$

where

$$
\begin{aligned}
a_p &= (2a + p - 1)/2 \\
b_p &= \tfrac{2b + S(\boldsymbol{\theta}, \boldsymbol{\gamma})}{2}.
\end{aligned} \tag{16}
$$

Now

$$
\begin{aligned}
p(\log(\widetilde{\boldsymbol{\lambda}}) \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}, \widetilde{\widetilde{\boldsymbol{\zeta}}}) &= \int\int\int p(\log(\widetilde{\boldsymbol{\lambda}}), \beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\zeta}}) \mathrm{d}\beta_0 \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \\
&= \int\int\int p(\log(\widetilde{\boldsymbol{\lambda}}) \mid \beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\zeta}}) p(\beta_0 \mid \boldsymbol{\lambda}, \sigma^2, \widetilde{\boldsymbol{\zeta}}) p(\boldsymbol{\beta} \mid \boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\zeta}}) p(\sigma^2 \mid \boldsymbol{\lambda}, \widetilde{\boldsymbol{\zeta}}) \mathrm{d}\beta_0 \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \\
&\propto \int\int\int (\sigma^2)^{-\widetilde{p}/2} \exp\Big\{ -\tfrac{1}{\sigma^2} \Big(\widetilde{\boldsymbol{\theta}} - \beta_0 \mathbf{1} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}\Big)^T \Big(\widetilde{\boldsymbol{\theta}} - \beta_0 \mathbf{1} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}\Big) \Big\} \\
&\cdot (\sigma^2)^{-1/2} \cdot \exp\Big\{ -\tfrac{p}{\sigma^2} \big(\beta_0 - \overline{\boldsymbol{\theta}}\big)^2 \Big\} \cdot \big|\sigma^2 \widehat{\boldsymbol{\Sigma}}\big|^{-1/2} \cdot \exp\Big\{ -\tfrac{1}{\sigma^2} \Big(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\Big)^T \widehat{\boldsymbol{\Sigma}}^{-1} \Big(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\Big) \Big\} \\
&\cdot \big(\sigma^2\big)^{-(a_p+1)} \cdot \exp\big\{ -b_p/\sigma^2 \big\} \mathrm{d}\beta_0 \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \\
&\propto \int\int\int \exp\Big\{ -\tfrac{1}{\sigma^2} \Big[\Big(\widetilde{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\mathbf{1} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}\Big)^T (I + \tfrac{1}{p}\mathbf{1}\mathbf{1}^T + \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{X}}^T)^{-1} \Big(\widetilde{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\mathbf{1} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}\Big) \\
&+ \Big(\boldsymbol{\beta} - \widehat{\widehat{\boldsymbol{\beta}}}\Big)^T \widehat{\widehat{\boldsymbol{\Sigma}}}^{-1} \Big(\boldsymbol{\beta} - \widehat{\widehat{\boldsymbol{\beta}}}\Big) + (p + \widetilde{p}) \Big(\beta_0 - \tfrac{1}{p+\widetilde{p}} \Big(p\overline{\boldsymbol{\theta}} + \widetilde{p}\overline{\widetilde{\boldsymbol{\theta}}}\Big)\Big)^2 + 2b_p \Big]\Big\} \cdot \big|\sigma^2 p(\widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}})^{-1}\big|^{-1/2} \\
&\cdot \big(\sigma^2\big)^{-(a_p+1+(\widetilde{p}+1)/2)} \mathrm{d}\beta_0 \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \\
&\propto \int \exp\Big\{ -\tfrac{1}{\sigma^2} \Big[\Big(\widetilde{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\mathbf{1} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}\Big)^T (I + \tfrac{1}{p}\mathbf{1}\mathbf{1}^T + \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{X}}^T)^{-1} \Big(\widetilde{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\mathbf{1} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}\Big) + 2b_p \Big]\Big\} \\
&\cdot \big|\sigma^2 p(\widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}})^{-1}\big|^{-1/2} \cdot \big(\sigma^2\big)^{-(a_p+1+\widetilde{p}/2)} \cdot \Big|\sigma^2 \widehat{\widehat{\boldsymbol{\Sigma}}}\Big|^{1/2} \mathrm{d}\sigma^2 \\
&\propto \Big(2b_p + \Big(\widetilde{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\mathbf{1} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}\Big)^T (I + \tfrac{1}{p}\mathbf{1}\mathbf{1}^T + \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{X}}^T)^{-1} \Big(\widetilde{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\mathbf{1} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}\Big)\Big)^{-(2a_p+\widetilde{t})/2} \\
&\propto \mathrm{MVT}_{\widetilde{p}}\big(2a_p, \log(\widetilde{\boldsymbol{A}}) + \overline{\boldsymbol{\theta}}\mathbf{1} + \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}, \tfrac{2b_p}{2a_p}\big(I + \tfrac{1}{p}\mathbf{1}\mathbf{1}^T + \tfrac{p}{p+1}\widetilde{\boldsymbol{X}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\widetilde{\boldsymbol{X}}^T\big)\big)
\end{aligned}
\tag{17}
$$

where $\widetilde{\boldsymbol{\theta}} = \log(\widetilde{\boldsymbol{\theta}}) - \log(\widetilde{\boldsymbol{A}})$,

$$
\begin{aligned}
\widehat{\widehat{\boldsymbol{\Sigma}}} &= \big(\widehat{\boldsymbol{\Sigma}}^{-1} + \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}}\big)^{-1} \\
\widehat{\widehat{\boldsymbol{\beta}}} &= \widehat{\widehat{\boldsymbol{\Sigma}}}\big(\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\beta}} + \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{\theta}}\big)
\end{aligned}
$$

and a_p and b_p are given in (16). When $r = 0$ and the model consists of intercept only, we get

$$
p(\log(\widetilde{\boldsymbol{\lambda}}) \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\zeta}}) = \mathrm{MVT}_{\widetilde{p}}\big(2a_p, \log(\widetilde{\boldsymbol{A}}) + \overline{\boldsymbol{\theta}}\mathbf{1}, \tfrac{2b_p}{2a_p}\big(I + \tfrac{1}{p}\mathbf{1}\mathbf{1}^T\big)\big).
$$

Fernandez *et al.* (2001) derived the analogue of (14) for predicting a single response in the Gaussian linear model setting.

A.4.1 Sampling from the posterior predictive distribution

The dimension of $\log(\widetilde{\boldsymbol{\lambda}})$ varies depending on $\widetilde{\boldsymbol{\zeta}}$, and hence $\widetilde{\boldsymbol{\zeta}}$ and $\log(\widetilde{\boldsymbol{\lambda}})$ should be sampled jointly. However, since we are sampling from a posterior predictive distribution (and hence avoid the "likelihood" term in the acceptance probability), a variant of Gibbs sampling with proposal

$$
\begin{aligned}
q((\widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\zeta}}), (\widetilde{\boldsymbol{\lambda}}', \widetilde{\boldsymbol{\zeta}}')) &= q(\widetilde{\boldsymbol{\zeta}}, \widetilde{\boldsymbol{\zeta}}') \cdot q(\widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\lambda}}') \\
&= p(\widetilde{\boldsymbol{\zeta}}' \mid \boldsymbol{\beta}^\gamma, \boldsymbol{\gamma}^\alpha) \cdot p(\log(\widetilde{\boldsymbol{\lambda}}') \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}^\lambda, \widetilde{\boldsymbol{\zeta}}')
\end{aligned}
$$

i.e. proposing $\widetilde{\boldsymbol{\zeta}}'$ conditioned on $\boldsymbol{\gamma}^\alpha$ and $\boldsymbol{\beta}^\gamma$, and then $\log(\widetilde{\boldsymbol{\lambda}})'$ conditioned on $\boldsymbol{\gamma}^\lambda$, $\boldsymbol{N}$ and $\widetilde{\boldsymbol{\zeta}}'$, which is sampling $(\widetilde{\boldsymbol{\zeta}}, \log(\widetilde{\boldsymbol{\lambda}}))$ directly from its joint "predictive full conditional", will have acceptance probability

$$
\min\left(1, \frac{p(\log(\widetilde{\boldsymbol{\lambda}}') \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}^\lambda, \widetilde{\boldsymbol{\zeta}}') p(\widetilde{\boldsymbol{\zeta}}' \mid \boldsymbol{\beta}^\gamma, \boldsymbol{\gamma}^\alpha) q((\widetilde{\boldsymbol{\lambda}}', \widetilde{\boldsymbol{\zeta}}'), (\widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\zeta}}))}{p(\log(\widetilde{\boldsymbol{\lambda}}) \mid \boldsymbol{\lambda}, \boldsymbol{\gamma}^\lambda, \widetilde{\boldsymbol{\zeta}}) p(\widetilde{\boldsymbol{\zeta}} \mid \boldsymbol{\beta}^\gamma, \boldsymbol{\gamma}^\alpha) q((\log(\widetilde{\boldsymbol{\lambda}}), \widetilde{\boldsymbol{\zeta}}), (\log(\widetilde{\boldsymbol{\lambda}})', \widetilde{\boldsymbol{\zeta}}'))}\right) = 1
$$

and hence it is easy to obtain posterior predictive samples within the existing posterior MCMC algorithm.

Acknowledgements

Research was funded by the centre Statistics for Innovation, (sfi)^2, the IAPP Marie Curie grant Climate Change and the Insurance Industry (CCII). This paper has greatly benefit from discussions with Magne Aldrin, Stein Beldring, Rasmus Benestad, Andrew Brown, Xeni Dimakos, Torill Engen-Skaugen, Eirik Førland, Peter Guttorp, Jan Erik Haugen, Lars Holden, Trond Iversen, Reason L. Machete, Hugo Maruri-Aguilar, Trevor Maynard, Falk Niehoerster, Elisabeth Nyeggen, Nicola Ranger, Håvard Rue, Kjersti Skeie, Lenny Smith and Henry Wynn. Map data kindly provided by the Norwegian Mapping Authority.

References

Association of British Insurers (2005). Financial risks of climate change. (Available from `http://www.abi.org.uk/Display/File/Child/552/Financial_Risks_of_Climate_Change.pdf`).

Bartlett, M. S. (1957). A comment on D. V. Lindley's statistical paradox. *Biometrika* **44**, 533–534.

Bottolo, L. & Richardson, S. (2008). Evolutionary stochastic search. *Tech. rep.* Available at `http://www.bgx.org.uk/publications.html`.

Botzen, W. J. & van den Bergh, J. C. (2008). Insurance against climate change and flooding in the Netherlands: Present, future, and comparison with other countries. *Risk Analysis* **28**, 413–426.

Boucher, J.-P. & Denuit, M. (2006). Fixed versus random effects in Poisson regression models for claim counts: A case study with motor insurance. *ASTIN Bulletin*.

Boucher, J.-P. & Guillen, M. (2009). A survey on models for panel count data with applications to insurance. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A, Matemáticas.* **103**, 277–294.

Chipman, H., George, E. I. & McCulloch, R. E. (2001). The practical implementation of Bayesian model selection (with discussion). In *Model Selection* (ed. P. Lahiri), 65–134. IMS: Beachwood, OH.

Fernandez, C., Ley, E. & Steel, M. F. J. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics* **100**, 381–427.

Frees, E. W., Shi, P. & Valdez, E. A. (2009). Actuarial applications of a hierarchical insurance claims model. *ASTIN Bulletin*.

Frees, E. W. & Valdez, E. A. (2008). Hierarchical insurance claims modeling. *J. Am. Statist. Ass.* **103**, 1457–1469.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing* **7**, 57–68.

George, E. I. & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339–373.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrica* **82**, 711–732.

Green, P. J. & Richardson, S. (2002). Hidden markov models and disease mapping. *J. Am. Statist. Ass.*

Haug, O., Dimakos, X. K., Vårdal, J. F. & Aldrin, M. (2008). Climate change and its impact on building water damage. `www.actuaries.org/ASTIN/Colloquia/Manchester/Papers/haug_paper_final.pdf`. 38th International ASTIN Colloquium.

Haug, O., Dimakos, X. K., Vårdal, J. F., Aldrin, M. & Meze-Hausken, E. (2009). Future building water loss projections posed by climate change. *Scandinavian Actuarial Journal*. DOI: 10.1080/03461230903266533.

Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* **36**, 531–547.

Kass, R. E. & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the schwartz criterion. *J. Am. Statist. Ass.* **90**, 928–934.

Kleindorfer, P. R. (2010). Interdependency of science and risk finance in catastrophe insurance and climate change. In *The Second International Conference on Asian Catastrophe Insurance, 8-9 December 2009*. Beijing, China.

Kohn, R., Smith, M. & Chan, D. (2001). Nonparametric regression using linear combinations of basis functions. *Statistics and Computing* **11**, 313–322.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–14.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models*. Chapman and Hall, London, 2nd edn.

Mills, E. (2005). Insurance in a climate of change. *Science* **309**, 1040–1044.

Mills, E., Roth, R. J. & Lecomte, E. (2005). Availability and affordability of insurance under climate change: A growing challenge for the U.S. Ceres Reports, www.ceres.org.

Muir-Wood, R., Miller, S. & Boissonade, A. (2006). The search for trends in a global catalogue of normalised weather-related catastrophe losses. Climate Change and Disaster Losses Workshop, Hohenkammer.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics* **33**, 341–365.

Nordhaus, W. (2008). *A Question of Balance*. Yale University Press.

Nott, D. J. & Leonte, D. (2004). Sampling schemes for Bayesian variable selection in generalized linear models. *Journal of Computational and Graphical Statistics* **13**, 362–382.

Ntzoufras, I., Dellaportas, P. & Forster, J. J. (2003). Bayesian variable and link determination for generalised linear models. *Journal of Statistical Planning and Inference* **111**, 165–180.

Smith, M. & Fahrmeir, L. (2007). Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *J. Am. Statist. Ass.* **102**, 417–431.

Smith, M. & Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics* **75**, 317–343.

Vellinga, P., Mills, E., Berz, G., Bouwer, L., Huq, S., Kozak, L. A., Palutikof, J., Schanzenbächer, B., Soler, G., Benson, C., Bruce, J., Frerks, G., Huyck, P., Kovacs, P., Olsthoom, A., Peara, A. & Shida, S. (2001). Insurance and other financial services. Chapter 8 of Climate Change 2001: Impacts, Adaptation and Vulnerability, Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change, `http://www.grida.no/climate/ipcc_tar/wg2/321.htm`.

Warner, K., Ranger, N., Surminski, S., Arnold, M., Linnnerooth-Bayer, J., Michel-Kerjan, E. & Paul Kovacs, C. H. (2009). Adaptation to climate change: Linking disaster risk reduction and insurance. *United Nations International Strategy for Disaster Reduction Secretariat (UNISDR)*.

West, M. (1985). Generalized linear models: outlier accommodation, scale parameters and prior distributions. In *Bayesian Statistics 2* (ed. J. M. B. *et al.*), 531–538. Amsterdam: North Holland.

Yip, K. C. H. & Yau, K. K. W. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics* **36**, 153–163.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian inference and decision techniques - Essays in honor of Bruno de Finetti* (eds. P. K. Goel & A. Zellner), 233–243. North-Holland, Amsterdam.