Fahrmeir, Lang:

Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data

Sonderforschungsbereich 386, Paper 202 (2000)

Online unter: http://epub.ub.uni-muenchen.de/

Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data

Ludwig Fahrmeir and Stefan Lang

University of Munich
Ludwigstrasse 33
D-80539 Munich

e-mail: fahrmeir@stat.uni-muenchen.de
lang@stat.uni-muenchen.de

SUMMARY

We present a unified semiparametric Bayesian approach based on Markov random field priors for analyzing the dependence of multicategorical response variables on time, space and further covariates. The general model extends dynamic, or state space, models for categorical time series and longitudinal data by including spatial effects as well as nonlinear effects of metrical covariates in flexible semiparametric form. Trend and seasonal components, different types of covariates and spatial effects are all treated within the same general framework by assigning appropriate priors with different forms and degrees of smoothness. Inference is fully Bayesian and uses MCMC techniques for posterior analysis. We provide two approaches: The first one is based on direct evaluation of observation likelihoods. The second one is based on latent semiparametric utility models and is particularly useful for probit models. The methods are illustrated by applications to unemployment data and a forest damage survey.

KEYWORDS: Categorical time-space data, forest damage, Markov random fields, MCMC, semiparametric Bayesian inference, unemployment.

1 Introduction

Multicategorical longitudinal data consists of observations (Y_{it}, x_{it}), $i = 1, \ldots, n$, $t = 1, \ldots, T$, for a population of n units observed across time, where the response variable Y is observed in ordered or unordered categories $r \in \{1, \ldots, k\}$. Covariates may be time-constant or time-varying. For T small compared to n, generalized estimating equation approaches are a popular choice for data analysis. For moderate or larger T, dynamic or state space models are a useful alternative, see, e.g., Fahrmeir and Tutz (1994, 2000, ch.8). For the special case (n=1) of categorical time series, dynamic generalized linear models are a meanwhile well established tool for approximate or full Bayesian inference.

In this paper, we consider multicategorical time-space data, where the spatial location or site s on a spatial array $\{1, \ldots, s, \ldots, S\}$ is given for each unit as an additional information. We also distinguish between metrical covariates $x_t = (x_{t1}, \dots, x_{tp})'$, whose effects will be modelled and estimated nonparametrically, and a further vector w_t of covariates, whose effects will be modelled parametrically in usual linear form. Multicategorical time-space data on n individuals or units then consists of observations

$$(Y_{it}, x_{it}, w_{it}, s_i), \quad i = 1, \dots, n, \quad t = 1, \dots, T, \tag{1.1}$$

where $s_i \in \{1, \dots, S\}$ is the location of individual i.

A typical example are monthly register data from the German Employment Office for the years 1980-1995, where Y_{it} is the employment status (e.g. unemployed, part time job, full time job) of individual i during month t and s_i is the district in Germany where i has its domicile. Data from surveys on forest health are a further example: Damage state Y_{it} of tree i in year t, indicated by the defoliation degree, is measured in ordered categories (severe to none) and s_i is the site of the tree on a lattice map. In both examples, covariates can be categorical or continuous, and possibly time-varying.

In general, time-space data of this kind cannot be analyzed adequately with existing nonparametric or conventional parametric methods. We present a unified semiparametric Bayesian framework for jointly modelling and analyzing effects of time, space and different types of covariates on categorical responses. Trend or seasonal components, spatial effects, metrical covariates with nonlinear effects and usual covariates with fixed effects are all treated within the same general framework by assigning appropriate priors with different forms and degrees of smoothness. This broad class of models contains state space models for categorical time series considered in previous work as a special case. Inference is fully Bayesian and uses recent MCMC techniques. We suggest two approaches for MCMC inference. The first one is useful if the likelihood of the data, given covariates and unknown parameters, can be easily computed as for cumulative or multinomial logistic models. Markov chain samples are then generated by an extension of Metropolis-Hastings algorithms developed in Fahrmeir and Lang (1999) for univariate responses. The second approach is based on latent variables, where the observable categorical responses are generated through threshold or utility mechanisms. For latent Gaussian variables this leads to multicategorical probit models, see Albert and Chib (1993) for the simpler case of linear predictors, and Yau, Kohn and Wong (2000) for nonparametric regression using basis functions. For MCMC inference, Gaussian latent variables are considered as unknown additional "parameters" and are generated jointly with the other parameters in a Gibbs sampling scheme. Efficient methods for sampling from high dimensional Gaussian Markov random fields are incorporated as a major building block.

Section 2 describes our Bayesian semiparametric regression models for categorical responses, observed across time and space, and depending on unknown functions and parameters. MCMC algorithms are presented in Section 3. In Section 4, the methods are applied to reemployment chances based on categorical time-space data on (un-) employment status and to data from a forest health inventory.

2 Semiparametric Bayesian models for multicategorical time-space data

Categorical response models may be motivated from the consideration of latent variables. This is not only useful for construction of models, but also for Bayesian inference, treating latent variables as additional unknown "parameters".
For the case of a *nominal response* Y with unordered categories 1,...,k, let U_r be a latent variable or utility associated with the r th category. Assume that U_r is given by

$$U_r = \eta_r + \varepsilon_r, \tag{2.1}$$

where η_r is a linear or semiparametric predictor depending on covariates and parameters, and $\varepsilon_1, \ldots, \varepsilon_k$ are random errors. Following the principle of random utility the observable response Y is determined by

$$Y = r \quad \Leftrightarrow \quad U_r = \max_{j=1,\ldots,k} U_j, \tag{2.2}$$

i.e., in choice situations the alternative is chosen which has maximal utility. Depending on the distributional assumptions for the error variables ϵ_r, equation (2.2) yields different models. If the ϵ's are i.i.d. normal, one gets the independent probit model. The more general multivariate probit model allows correlated noise variables. Assuming i.i.d. error variables following the extreme value distribution $F(z) = exp(-exp(-z))$ yields the multinomial logit model

$$P(Y = r) = exp(\eta_r)/(exp(\eta_1) + \ldots + exp(\eta_k)).$$

Since only differences of utilities are identifiable we may set $\eta_k = 0$ for the reference category k. With a linear predictor $\eta_r = w'\beta_r$, one obtains the common form

$$P(Y = r) = exp(\eta_r)/(1 + exp(\eta_1) + \ldots + exp(\eta_{k-1})), \quad r = 1, \ldots, k-1,$$

of a parametric multinomial logit model.
For the case of an *ordered response* Y, cumulative models based on a threshold approach are most widely used. It is postulated that Y is a categorized version of a latent variable

$$U = \eta + \varepsilon, \tag{2.3}$$

obtained through the threshold mechanism

$$Y = r \Leftrightarrow \theta_{r-1} < U \leq \theta_r, \quad r = 1, \ldots, k, \tag{2.4}$$

with thresholds $-\infty = \theta_0 < \theta_1 < \ldots < \theta_k = \infty$. If the error variable ε has distribution function F, it follows that Y obeys a cumulative model

$$P(Y \leq r) = F(\theta_r - \eta). \tag{2.5}$$

With a linear predictor $\eta = w'\beta$ one gets parametric cumulative models. For identifiability reasons, the linear combination does not contain an intercept term β_0. Otherwise one of the thresholds, for example θ_1, had to be set to zero.
The most popular choices for F in (2.5) are the logistic and the standard normal

distribution function leading to cumulative logit or probit models.

For multicategorical time-space data (1.1), we generally assume more flexible semiparametric predictors. For nominal responses Y_{it}, the general form of a semiparametric additive predictor associated with category r is

$$\eta_{itr} = f^r_{time}(t) + f^r_{spat}(s_i) + \sum_{j=1}^{p} f^r_j(x_{itj}) + w'_{it}\beta_r. \tag{2.6}$$

Here f^r_{time} and f^r_{spat} represent possibly nonlinear effects of time and space, $f^r_1, \ldots, f^r_p$ are unknown smooth functions of the metrical covariates $x_1, \ldots, x_p$, and $w'_{it}\beta_r$ corresponds to the usual parametric part of the predictor. Note that the latter may also contain a (category specific) intercept. Depending on the analysed dataset, the effect of time f^r_{time} may contain only a nonlinear time trend or may be split up into a trend and a seasonal component, i.e.

$$f^r_{time}(t) = f^r_{trend}(t) + f^r_{season}(t).$$

In analogy we usually split up the spatial effect f^r_{spat} into a spatially correlated (structured) and an uncorrelated (unstructured) effect

$$f^r_{spat}(s) = f^r_{str}(s) + f^r_{unstr}(s)$$

A rational therefore is that a spatial effect is usually a surrogate of many unobserved influential factors, some of them may obey a strong spatial structure and others may be present only locally. By estimating a structured and an unstructured effect we aim at separating between the two kinds of influential factors. As a side effect we are able to assess to some extent the amount of spatial dependency in the data by observing which one of the two effects exceeds. If the unstructured effect exceeds, the spatial dependency is smaller and vice versa. With the same arguments we could also divide up the time trend $f^r_{Trend}(t)$ into a correlated and an uncorrelated component. Such models are common in spatial epidemiology, see Besag, York, Mollie (1991) and Knorr-Held, Besag (1998).

A further extension of (2.6) are varying coefficient models, where nonlinear terms $f^r_j(x_{itj})$ are generalized to $f^r_j(x_{itj})z_{itj}$, where z_j may be a component of x or w or a further covariate. Covariate x_j is called the effect modifier of z_j because the effect of z_j varies smoothly over the range of x_j. Of course, time t and even the spatial covariate s are also possible effect modifiers.

Finally, we note that the effect of a particular covariate in (2.6) may be present only for some of the k categories of the response. In addition, we may observe covariates that are associated only with one specific category of the response, so called category specific covariates. This leads to the distinction between category specific and global covariates. Although a modification of (2.6) and estimation of such models is straightforward (and already implemented) we do not go into details here, because the inclusion of category specific covariates is not necessary for the applications of this paper. More details can be found in Fahrmeir, Tutz (1994, 2000, Ch. 3) and the references therein.

For ordered responses Y_{it} following a cumulative model (2.5), we assume semiparametric predictors

$$\eta_{it} = f_{time}(t) + f_{spat}(s_i) + \sum_{j=1}^{p} f_j(x_{itj}) + w'_{it}\beta. \tag{2.7}$$

where the terms have the same interpretation as in (2.6), omitting the category-specific index r. Note that the term $w'\beta$ for fixed effects must not contain an intercept to make thresholds identifiable.

For Bayesian inference, unknown functions f_{time},f_{spat},$f_1, \dots, f_p$, thresholds $\theta = (\theta_1, \dots, \theta_{k-1})'$ and all other parameters are considered as random variables. Categorical response models are to be understood conditional upon these random variables and have to be supplemented by appropriate prior distributions. For the "fixed effect" parameters θ and β we assume diffuse priors

$$p(\theta) \propto \text{const}, \quad p(\beta) \propto \text{const}.$$

Priors for a time trend f_{trend} of time t and functions $f_1, \dots, f_p$ of metrical covariates are specified by local smoothness priors common in state space modelling of structural time series. We illustrate the approach for the effect of a specific metrical covariate x. Let

$$x_{(1)} < \dots < x_{(l)} < \dots < x_{(m)},$$

denote the m *different, ordered* observed values of the metrical covariate x. Define $f(l) := f(x_{(l)})$ and let

$$f = (f(1), \dots, f(l), \dots, f(m))'$$

denote the corresponding vector of function evaluations. For equally-spaced values $x_{(1)}, \dots, x_{(m)}$ we usually assign first or second order random walk models

$$f(l) = f(l-1) + \xi(l) \quad \text{or} \quad f(l) = 2f(l-1) - f(l-2) + \xi(l) \tag{2.8}$$

with Gaussian errors $\xi(l) \sim N(0; \tau^2)$ and diffuse priors $f(1) \propto$ const, or $f(1)$ and $f(2) \propto$ const, for initial values, respectively. Both specifications act as smoothness priors that penalize too rough functions f. The variance τ^2 controls the degree of smoothness of $f = (f(1), \dots, f(l), \dots, f(m))'$. Of course, local linear trend models or higher order autoregressive priors are also possible. An example is a time varying seasonal component f_{Season} of time t. A flexible seasonal component with period *per* can be defined by

$$f_{season}(t) = -\sum_{j=1}^{per-1} f_{season}(t-j) + \xi(t) \tag{2.9}$$

and once again diffuse priors for initial values.

For non-equally spaced values $x_{(1)}, \dots, x_{(m)}$, priors have to be modified to account for nonequal distances $\delta_l = x_{(l)} - x_{(l-1)}$. Random walks of first order are now specified by

$$f(l) = f(l-1) + \xi(l), \quad \xi(l) \sim N(0, \delta_l \tau^2),$$

and random walks of second order by

$$f(l) = (1 + \frac{\delta_l}{\delta_{l-1}})f(l-1) - \frac{\delta_l}{\delta_{l-1}} f(l-2) + \xi(l), \quad \xi(l) \sim N(0, \gamma_l \tau^2)$$

with appropriate weights γ_l. Based on Fahrmeir, Lang (2000) we choose $\gamma_l = \delta_l(1 + \frac{\delta_l}{\delta_{l-1}})$.

All these priors can be equivalently rewritten in form of a global smoothness prior

$$f|\tau^2 \propto exp\left(-\frac{1}{2\tau^2} f' K f\right), \tag{2.10}$$

with appropriate penalty matrix K. For example,

$$K = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}$$

in the simple case of a first order random walk and equidistant observations.

Let us now turn our attention to a spatial covariate s, where the values of s represent the location or site in geographical regions. For the spatially correlated effect $f_{str}(s)$, $s = 1, \ldots, S$, we choose Markov random field priors common in spatial statistics (Besag, York and Mollie, 1991). These priors reflect spatial neighbourhood relationships. For geographical data one usually assumes that two sites or regions s_i and s_j are neighbours if they share a common boundary. Then a spatial extension of random walk models leads to the conditional, spatially autoregressive specification

$$f_{str}(s)|f_{str}(u), u \neq s \sim N\left(\sum_{u \in \partial_s} \frac{1}{N_s} f_{str}(u), \frac{\tau^2_{str}}{N_s}\right), \tag{2.11}$$

where N_s is the number of adjacent regions, and $u \in \partial_s$ denotes that region u is a neighbour of region s. Thus the (conditional) mean of $f_{str}(s)$ is an average of function evaluations $f_{str}(u)$ of neighboring regions. Again the variance τ^2_{str} controls the degree of smoothness. This prior will be used in our first application on durations of unemployment. In some applications, as in our second example on forest damage data, a more general prior specification seems to be more appropriate. In this application we assume that two sites s_i and s_j are neighbors if they are within a certain distance, d say. In addition, we assume that the conditional mean of $f_{str}(s)$ is now a weighted average of function evaluations $f_{str}(u)$ of neighboring sites rather than an unweighted average as in (2.11). The weights are chosen to be proportional to the distance of neighboring sites to site s. In terms of weights w_{su} a general spatial prior can be defined as

$$f_{str}(s)|f_{str}(u), u \neq s \sim N\left(\sum_{u \in \partial_s} \frac{w_{su}}{w_{s+}} f_{str}(u), \frac{\tau^2_{str}}{w_{s+}}\right), \tag{2.12}$$

where $+$ denotes summation over the missing subscript. In the forest damage apliaction the weights w_{su} are equal to the distance of site s and u. Note that the spatial prior (2.11) is a special case of (2.12) with weights $w_{su} = 1$.

As for autoregressive priors, (2.12) can be written in the form (2.10), where the elements of the penalty matrix K are given by

$$k_{ss} = w_{s+}$$

and

$$k_{su} = \begin{cases} -w_{su} & u \in \partial_s \\ 0 & else. \end{cases}$$

As mentioned before, we usually split up the effect of a spatial covariate into a structured (spatially correlated) and an unstructured (uncorrelated) effect. For the unstructured effect f_{unstr} a common assumption is that the parameters $f_{unstr}(s)$ are i.i.d. Gaussian

$$f_{unstr}(s)|\tau^2_{unstr} \sim N(0, \tau^2_{unstr}). \tag{2.13}$$

Note that we are not restricted to an unstructured effect only for the spatial covariate s. An unstructured effect for time t, or with respect to any other grouping variable, is also possible (and already supported in our implementation).

For a fully Bayesian analysis, variance or smoothness parameters τ_j^2, $j = trend, season, str, unstr, 1, \ldots, p$ are also considered as unknown and estimated simultaneously together with unknown functions. Therefore, hyperpriors are assigned to them in a second stage of the hierarchy by highly dispersed inverse gamma distributions

$$p(\tau_j^2) \sim IG(a_j, b_j)$$

with known hyperparameters a_j and b_j. It turns out that the simultaneous estimation of smooth functions and smoothing parameters is a great advantage of our Bayesian modelling approach. In a frequentist approach smoothing parameters are usually chosen by minimizing some goodness of fit criteria (e.g. AIC) with respect to the smoothing parameters, or via cross validation. However, if the model contains many nonparametric effects as in the applications of this paper, a multidimensional grid search is required which becomes totally inpractical for higher dimensions. This problem gets even worse in multicategorical response models.

The Bayesian model is completed by the following conditional independence assumptions:

(i) For given covariates and parameters observations Y_{it} are conditionally independent.

(ii) Priors for function evaluations, fixed effects parameters and for variances are all mutually independent.

3 Posterior analysis via MCMC

In the following f denotes the vector of all function evaluations including trend and seasonal components of time t and structured and unstructured spatial effects, τ is the vector of all variances, $\gamma = \beta$ for nominal and $\gamma = (\beta, \theta)$ for ordinal models. For a nominal logit model or a cumulative logit model, the contribution of Y_{it} to the likelihood $p(Y|f, \gamma)$ of the data given the parameters can be easily calculated. Bayesian inference can then be based on the posterior

$$p(f, \tau, \gamma|Y) \propto p(Y|f, \gamma)p(f|\tau)p(\tau)p(\gamma).$$

MCMC simulation is based on drawings from full conditionals of single parameters or blocks of parameters, given the rest and the data. Single moves update each parameter separately. Convergence and mixing is considerably improved by block moves for the vectors $f_j = (\ldots, f_j(l), \ldots)'$ of function evaluations, where blocks $f_j[u, v] = (f_j(u), \ldots, f_j(v))'$ of parameters are updated instead of single parameters $f_j(l)$.

Markov chain samples for $f_j[u,v]$ from the unnormalized full conditionals $p(f_j[u,v]|\cdot)$ are generated by Metropolis-Hastings (MH) steps with conditional prior proposals as suggested by Knorr-Held (1999). Drawings from $p(\gamma|\cdot)$ and the unstructured spatial effect f_{unstr} (or other unstructured effects) can be obtained by the weighted least squares proposal of Gamerman (1997) or a slight modification. Updating of variance parameters is done by Gibbs steps, drawing directly from inverse gamma densities. Details of the updating schemes are described in Fahrmeir and Lang (1999) for univariate responses.
The resulting hybrid MCMC scheme for generating posterior samples is then defined by drawing from the following full conditionals:

Sampling scheme 1:

(i) Draw samples for "fixed effects" parameters γ by MH steps with weighted least squares proposals. For cumulative models (2.5), thresholds have to obey order restrictions.

(ii) For each function f, partition the vector of function evaluations into blocks $f^{(b)}$,$b = 1, 2, ...$, and draw from

$$p(f^{(b)}|\cdot), \quad b = 1, 2, \ldots,$$

with MH steps using conditional prior proposals.

(iii) Draw samples for unstructured spatial effects $f_{unstr}(s)$, $s = 1, \ldots, S$ by MH steps with weighted least squares proposals.

(iv) Draw samples for variances τ_j^2 from inverse Gamma posteriors

$$p(\tau_j^2|\cdot) \sim IG(a_j', b_j')$$

with updated parameters a_j', b_j' given by $a_j' = a_j + \frac{rank(K_j)}{2}$ and $b_j' = b_j + \frac{1}{2} f_j' K_j f_j$.

For categorical responses, useful alternative sampling schemes can be developed on the basis of the latent variable mechanisms (2.2) and (2.4), augmenting the observables Y_{it} by corresponding latent variables

$$U_{itr} = \eta_{itr} + \epsilon_{itr} \quad \text{or} \quad U_{it} = \eta_{it} + \epsilon_{it},$$

respectively, with semiparametric predictors as in (2.6) or (2.7). Assuming Gaussian errors, we obtain multicategorical probit models with latent semiparametric Gaussian models. Posterior analysis is now based on

$$p(f, \gamma, \tau, U|Y) \propto p(Y|U)p(U|f, \gamma)p(f|\tau)p(\tau)p(\gamma),$$

with $p(Y|U) = \prod_{i,t} p(Y_{it}|U_{it})$, where $U_{it} = (U_{it1}, \ldots, U_{itk})'$ for nominal responses. The conditional likelihood $p(Y_{it}|U_{it})$ is determined by the mechanisms (2.2) or (2.4). For a nominal response, we have

$$p(Y_{it}|U_{it}) = \sum_{r=1}^{k} I(max(U_{it1}, \ldots, U_{itk}) = U_{itr})I(Y_{it} = r). \qquad (3.1)$$

For a cumulative model, we get

$$p(Y_{it}|U_{it}) = \sum_{r=1}^{k} I(\theta_{r-1} < U_{it} \leq \theta_r) I(Y_{it} = r), \tag{3.2}$$

due to the fact that $p(Y_{it}|U_{it})$ is one if U_{it} obeys the constraint imposed by the observed value of Y_{it}. Compared to the direct sampling scheme above, additional drawings from full conditionals for the latent variables U_{it} are necessary. As an advantage, full conditionals for functions and fixed effects parameters become Gaussian, allowing computationally efficient Gibbs sampling. The full conditionals for U_{it} are:

$$p(U_{it}|f, \gamma, Y_{it}) \propto p(Y_{it}|U_{it}) p(U_{it}|f, \gamma). \tag{3.3}$$

Since latent variables U_{it} have (conditional) Gaussian distributions with means η_{it} and unit variances, their full conditionals are truncated standard normals, with truncation points determined by the restrictions (3.2) and (3.1).

To derive full conditionals for functions f_j and fixed effects parameters β it is convenient to rewrite the predictors (2.6) and (2.7) in matrix notation. For example for (2.7) we obtain

$$\eta = X_{time} f_{time} + X_{spat} f_{spat} + \sum_{j=1}^{p} X_j f_j + W\beta. \tag{3.4}$$

Here the X_j are 0/1 matrices where the number of columns is equal to the number of parameters of the respective effect. If for observation i, t the value of covariate x_j (or time t or site s) is l, then the element in the i, t th row and the l th column is one, zero otherwise. Now standard calculations show that the full conditional for a function f_j is Gaussian with covariance matrix

$$\Sigma_j = P_j^{-1} = (X_j' X_j + \frac{1}{\tau_j^2} K_j)^{-1} \tag{3.5}$$

and mean

$$\mu_j = (X_j' X_j + \frac{1}{\tau_j^2} K_j)^{-1} X_j' (U - \tilde{\eta}), \tag{3.6}$$

where $\tilde{\eta}$ is the part of the predictor associated with all remaining effects in the model. Since $X_j' X_j$ is diagonal and the penalty matrix K_j is a bandmatrix (e.g. with bandwidth two for a second order random walk) it follows that the posterior precision P_j is also a bandmatrix with the same bandwith. Following Rue (2000), drawing random numbers from the full conditionals for f_j is as follows:

(i) Compute the Cholesky decomposition $P_j = L'L$.

(ii) Solve $L f_j = z$, where z is a vector of independent standard Gaussians. It follows that $f_j \sim N(0, \Sigma_j)$

(iii) Compute the mean μ_j by solving $P_j \mu_j = X_j'(U - \tilde{\eta})$. This is achieved by first solving by forward substitution $L'\nu = X_j'(U - \tilde{\eta})$ followed by backward substitution $L\mu_j = \nu$.

(iv) Set $f_j = f_j + \mu_j$, then $f_j \sim N(\mu_j, \Sigma_j)$.

All algorithms involved take advantage of the bandmatrix structure of the posterior precision P_j.

Finally, the full conditionals for fixed effects parameters β with diffuse priors are Gaussian with mean and covariance matrix given by

$$\mu_{beta} = (W'W)^{-1}W'(U - \tilde{\eta}), \qquad \Sigma_{beta} = (W'W)^{-1}. \tag{3.7}$$

We can now summarize the resulting sampling schemes. For a cumulative probit model, a Gibbs sampling scheme is defined by the following steps.

Sampling scheme 2:

(i) The latent variables $U_{it}, i = 1, \ldots, n, t = 1, \ldots, T$ are sampled as follows. If $Y_{it} = r$, then U_{it} is generated from $N(\eta_{it}, 1)$, with mean η_{it} as in (2.7), evaluated at current values of f_j and β, subject to the constraint $\theta_{r-1} < U_{it} \leq \theta_r$.

(ii) Following Albert and Chib (1993), the full conditional for threshold $\gamma_r, r = 1, \ldots, k-1$ is uniform on the interval

$$[max\{U_{it} : Y_{it} = r\}, min\{U_{it} : Y_{it} = r + 1\}].$$

Posterior samples from these uniform distribution may exhibit bad mixing. A reason is that intervals can become quite small and, as a consequence, the chain moves slowly. In such a case, other parametrizations as suggested for example in Chen and Dey (2000) are a possible alternative. For $k = 3$ such a reparametrization becomes particularly convenient, see our application to forest damage in Section 4.

(iii) Function evaluations f_j are generated from Gaussian full conditionals $p(f_j|U, \cdot)$ with covariance matrix (3.5) and mean (3.6), using the algorithms for bandmatrices described above.

(iv) Samples for variances are generated from inverse Gamma priors with updated parameters given in sampling scheme 1.

(v) Samples for fixed effects β are drawn from Gaussian full conditionals with mean and covariance matrix in (3.7).

For nominal response, we choose k as the reference category. Since only differences of utilities can be identified (see Section 3.2), we may either set the predictor η_{itk} to zero or the latent variable U_{itk}.

Sampling scheme 3:

(i) Setting $U_{itk} \equiv 0$, latent variables $U_{itr}, r = 1, \ldots, k-1$, are generated as follows for each observation $Y_{it}, i = 1, \ldots, n,\ t = 1, \ldots, T$.
If $Y_{it} = r,\ r \neq k$, then U_{itr} is generated first from a normal distribution with

mean η_{itr} and variance 1, subject to the constraints $U_{itr} > U_{itl}$, $l \neq k$, and $U_{itr} > 0$ ($\equiv U_{itk}$). Next we generate U_{itl} for $l \neq r$ from a normal distribution with mean η_{itl} and variance 1, subject to the constraint that U_{itl} is less than the U_{itr} generated just before.
If $Y_{it} = k$ (the reference category), then we generate $U_{itl}, l = 1, \ldots, k-1$, from a normal with mean η_{itl} and variance 1, subject to the constraint $U_{itl} < 0$.

(ii) Posterior samples for functions f_j^r, $r = 1, \ldots, k-1$ and all other parameters are generated as in the steps (iii)-(v) of sampling scheme 2.

4 Applications

We consider two applications. In a first application on unemployment durations we analyse unemployment data from the German Federal Employment Office ("Bundesanstalt für Arbeit"). This is a huge dataset with approximately 280000 observations showing the practicability of our methods even for very large datasets. In a second application, we analyse longitudinal data on forest health collected in the forest district of Rothenbuch in northern Bavaria for the years 1983-1997. All computations have been carried out with *BayesX*, a software package for Bayesian inference that has been developed at our department. The program is available for public use under `http://www.stat.uni-muenchen.de/~lang/`. See also Lang, Brezger (2000) for a detailed description of the capabilities of *BayesX*.

4.1 Reemployment chances

In our first application we analyse monthly unemployment data from the German Federal Employment Office for the years 1980-1995. Our analysis is restricted to data from former West Germany (excluding Berlin) and to women. For each individual the data provides information about the employment status in month t, the district where the individual lives and a number of personal characteristics. Since we are interested in analyzing reemployment chances, distinguishing between full and part time jobs, we define three-categorical response variables Y_{it} as event indicators

$$Y_{it} = \begin{cases} 1, \text{ gets a new full time job in month } t \text{ (calendar time)} \\ 2, \text{ gets a new part time job in month } t \\ 3, i \text{ is unemployed in month } t \text{ (reference category).} \end{cases}$$

Our analysis is based on the following covariates:

D	duration time measured in months
A	age (in years) at the beginning of unemployment
N	nationality, dichotomous with categories "german" and "foreigner" (= reference category)
U_d	unemployment compensation (in month d of duration time), dichotomous with categories "unemployment benefit" (=reference category) and "unemployment assistance"
P_t	number of previous unemployment periods (in month t of calendar time): 1,2,3 and more, 0 (reference category)

E education, trichotomous with categories "no vocational training" "vocational training" (reference category) and university
S district in which the unemployed have their domicil

All categorical covariates are coded in effect coding.
Then we model the probabilities $P(Y_{it} = r|\eta_{itr})$, $r = 1, 2$, by an independent probit model, with predictors

$$\begin{aligned}\eta_{itr} &= f^r_{trend}(t) + f^r_{season}(t) + f^r_{str}(S_i) + f^r_{unstr}(S_i) + \\ &\quad f^r_1(D_{it}) + f^r_2(A_{it}) + w'_{it}\beta_r, \quad r = 1, 2,\end{aligned}$$

where f^r_{trend} and f^r_{season} are trend and seasonal component of calendar time t, f^r_{str} and f^r_{unstr} are structured and unstructured spatial effects of the district, f^r_1 is the effect of duration D in current unemployment status and f^r_2 is the effect of age A. The priors for f^r_{trend}, f^r_1 and f^r_2 are second order random walk models (2.8). For f^r_{str} and f^r_{unstr} we assign the Markov random field prior (2.11) and the exchangeable prior (2.13), respectively. For the seasonal component we choose the flexible seasonal prior (2.9). Priors for fixed effects parameters β_r are diffuse. An analysis with similar predictors using a multinomial logit model and sampling scheme 1 for drawing samples from full conditionals can be found in Fahrmeir, Lang (2000).

Figure 2 displays estimated effects of duration time and calendar time trend and seasonal component for getting full time jobs (left column) and part time jobs (right column). Duration time effects have the typical pattern also observed in other investigations, with a peak after 2-3 months and sloping downward then. Calendar time trends for full and part time jobs show a similar general pattern: declining until the year 1982, then slowly increasing until 1990 (one year after the German reunion), declining distinctly again thereafter, with an intermediate recovery. This corresponds to the observed economic trend of the labor market in Germany during this period. Estimated seasonal effects are more or less stable over this period although varying in size. To gain more insight Figures f) and g) displays a section of the estimated effects for the year 1992 with the typical peaks in spring and autumn, and a global minimum in Dezember. The effect of age can be found in Figure 1. For the age effect there are local minima for women about 30, which may be a "family" effect. The dramatic decline of unemployment probabilities of people older than 50 years is particularly striking. The increase after 60 may be caused by boundary effects and we do not interpret it. Note that the age effect is much stronger for women seeking full time jobs (Figure a)) compared to women seeking part time jobs (Figure b)). Structured regional effects are shown in Figures 3 and 4. Figure 3 shows the estimated posterior mean and Figure 4 shows "probability" maps where the levels correspond to "significantly negative" (black colored), "nonsignificant" (grey colored), i.e. zero is within the confidence interval around the estimate, and "significantly positive" (white colored). In order to interpret the structured effects, unstructured effects must be taken into consideration as well. Therefore Table 4 gives a summary of the estimated posterior means of the unstructured spatial effect for the different regions. We observe that the structured effect for getting full time jobs is stronger than for getting part time jobs. Even more important, the unstructured effect for part time jobs cleary exceeds the structured effect which is in constrast to the estimated effects for full time jobs. Although the estimated posterior mean of the structured effect for getting part time jobs in Figure 3 b) shows some spatial variation, Figure 4 cleary indicates that there is no "significant" variation in terms of posterior probabilities. In the contrary, the structured effect for getting full time jobs displays "significant" variation with improved chances in the south compared to the middle and the north. The two dark spots in Figures 3 a) and 4 b) mark areas that are known for their structural economical problems during the eighties and nineties.

Estimates of fixed effects for getting full time jobs are shown in Table 2 and for getting part time jobs in Table 3. Tables 2 and 3 confirm some facts already known from previous analyses with more conventional methods. Chances for re-employment are better for Germans and for women with a university degree compared to women with vocational training and no vocational training. Both effects are stronger for women getting part time jobs. The number of previous unemployment periods serves as a surrogate for experience at the labor market: an increase in the number of previous spells increases the probability for shorter unemployment duration. The estimated effect of unemployment assistance is significantly negative and positive for umenployment benefits, which seems to contradict the widely-held conjecture about negative side-effects of unemployment benefits. However, it may be that the variable "unemployment benefit" also acts as a surrogate variable for those who have worked, and therefore contributed regularly to the insurance system in the past.

4.2 Forest health

In this longitudinal study on the state of trees, we analyse the influence of calendar time, age of trees, canopy density and location of the stand on the defoliation degree of beeches. Data have been collected in yearly forest damage inventories carried out in the forest district of Rothenbuch in northern Bavaria from 1983 to 1997. There are 80 observation points with occurence of beeches spread over an area extending about 15 km from east to west and 10 km from north to south, see Figure 5. The degree of defoliation is used as an indicator for the state of a tree. It is measured in three ordered categories, with $Y_{it} = 1$ for "bad" state of tree i in year t, $Y_{it} = 2$ for "medium" and $Y_{it} = 3$ for "good". A detailed data description can be found in Göttlein and Pruscha (1996). Covariates used here are defined as follows:

- A age of tree at the beginning of the study in 1983, measured in three effect coded categories a_1 = "below 50 years", a_2 = between 50 and 120 years, and a_3 = above 120 years (reference category);
- C Canopy density at the stand measured in percentages 0%,10%,...,90%,100%.

The covariate age is time constant by definition, while canopy density is time varying. Based on previous analysis, we use a three-categorical ordered probit model (2.5) based on a latent semiparametric model $U_{it} = \eta_{it} + \epsilon_{it}$ with predictor

$$\eta_{it} = f_{trend}(t) + f_{str}(s_i) + f(c_{it}) + \beta_1 a_{i1} + \beta_2 a_{i2}. \tag{4.1}$$

Here a_{i1} and a_{i2} are the indicators for age categories 1 and 2. The calendar time trend $f_{trend}(t)$ and the effect $f(c)$ of canopy density are modelled by random walks

Variable	mean	Std.Dev.	10% quant.	median	90% quant.
const	1.102	0.073	1.008	1.099	1.195
a_1	0.358	0.107	0.222	0.356	0.490
a_2	0.010	0.080	-0.092	0.009	0.113
a_3	-0.3677	0.085	-0.473	-0.369	-0.257

Table 1: Estimates of constant parameters

of second order. For the structured spatial effect we assign the Markov random field prior (2.12), with the neighborhood ∂_s of trees including all trees u with euclidian distance $d(s,u) \leq 1.2$ km and with weights defined by $w_{su} = d(s,u)$. An unstructured spatial effect is excluded from the predictor for the following two reasons. First, a look at the map of observation points (Figure 5) reveals some sites with only one neighbor, making the identification of a structured and an unstructured effect difficult if not impossible. The second reason is that for each of the 80 sites only 15 observations on the same tree are available with only minor changes of the response category. In fact, there are only a couple of sites were all three response categories have been observed. Thus, the inclusion of an unstructured effect in our model leads to severe inditification problems between the structured and unstructured effect, which can be observed by inspecting sampling paths of parameters.

We first applied the ordered probit model in standard parametrization. However, in step (ii) of sampling scheme 2 mixing of posterior samples for thresholds θ_1 and θ_2 was not satisfactory, see Figure 6. Following Chen and Dey (2000), we therefore reparametrized the model. First, inclusion of a constant β_0 in (4.1) allows to set $\theta_1 = 0$. Secondly, because parameters in the predictor of the latent Gaussian model are only identifiable up to a multiplicative factor, we assume that errors ϵ_{it} are $N(0, \sigma^2)$ distributed with unknown variance σ^2. This allows us to set $\theta_1 = 1$. The parameter β_0 is sampled simultaneously with fixed effects β_1 and β_2. For σ^2 we specify an inverse Gamma prior, leading to posterior samples from an inverse Gamma full conditional.

For interpretation of estimation results note the following: In accordance with our definitions (2.3) to (2.5), higher (lower) values of the predictor (4.1) (or of effects in this predictor) correspond to healthier (worse) state of the trees. Estimates for β_0 and the effect of age are given in Table 1. As we might have expected younger trees are in healthier state than the older ones. Figure 7 shows posterior mean estimates for the calendar time trend and for the effect of canopy density. We see that trees recover after the bad years around 1986, but after 1994 health status declines to a lower level again. The distinct monotonic increase of the effect of canopy densities $\geq$ 30% gives evidence that beeches get more shelter from bad environmental influences in stands with high canopy density. Figure 8 shows the estimated (structured) spatial effect in form of posterior probabilities, were black spots indicate areas with strictly negative credible regions, i.e. areas with more trees in bad state. The black colored sites correspond mostly to areas in the forest district which are located higher above sea level than the other sites. Here the environmental conditions in terms of nutrient quantity and soil quality are worse compared to other areas.

5 Conclusion

The applications demonstrate that the Bayesian methods developed are useful and flexible tools for inference in realistically complex categorical regression models.

A variety of extensions are possible by modifying or generalizing the observation models, predictors and smoothness priors. For example, probit models based on latent utilities can be extended to correlated categorical or mixed continous-categorical responses by considering latent multivariate semiparametric Gaussian models. Predictors can be made more flexible by introducing nonparametric interactions between covariates following suggestions in Clayton (1996) and Knorr-Held (2000). Replacing Gaussian priors by heavy-tail distributions would allow to consider unsmooth regression functions.

References

Albert, J., Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Ass.*, **88**, 669-679.

Besag, J., York, J., Mollie, A. (1991) Bayesian image restoration with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.*, **43**, 1-59.

Chen, M.H., Dey, D.K. (2000): Bayesian Analysis for Correlated Ordinal Data Models. In: Dey, D.K., Ghosh, S.K. and Mallick, B.K. (eds.), *Generalized linear models: A Bayesian perspective.* New York: Marcel Dekker.

Clayton, D. (1996) Generalized linear mixed models. In: Gilks, W., Richardson S. and Spiegelhalter D. (eds.), *Markov Chain Monte Carlo in Practice.* London: Chapman and Hall, 275-301.

Fahrmeir, L., Tutz, G. (1994, 2nd ed.2000) *Multivariate Statistical Modelling based on Generalized Linear Models.* New York: Springer–Verlag.

Fahrmeir, L., Lang, S. (1999): Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors. Discussion Paper 169, Sonderforschungsbereich 386, Ludwigs-Maximilians-Universität München. Available under `http://www.stat.uni-muenchen.de/sfb386/publikation.html`.

Fahrmeir, L., Lang, S. (2000): Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data. Proceedings of the International Syposium on Frontiers of Time Series Modeling in Tokio. Available under `http://www.stat.uni-muenchen.de/~lang/`.

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

Göttlein , A., Pruscha, H. (1996): Der Einfluss von Bestandskenngrössen, Topographie, Standord und Witterung auf die Entwicklung des Kronenzustandes im Bereich des Forstamtes Rothenbuch. *Forstwissenschaftliches Centralblatt*, **114**, 146-162.

Yau, P. Kohn, R., Wood, S. (2000): Bayesian Variable Selection and Model Averaging in High Dimensional Multinomial Nonparametric Regression. Preprint, University of New South Wales.

Knorr-Held, L. (1999): Conditional Prior Proposals in Dynamic Models. *Scand. J. Statist.*, **26**, 129-144.

Knorr-Held, L. (2000): Bayesian Modelling of Inseparable Space-Time Variation in Disease Risk. *Statistics in Medicine*, to appear.

Knorr-Held, L., Besag, J. (1998) Modelling Risk from a Desease in Time and Space. *Statistics and Medicine*, **17**, 2045-2060.

Lang, S., Brezger, A. (2000) BayesX - Software for Bayesian Inference based on Markov Chain Monte Carlo simulation techniques. Discussion Paper 187, Sonderforschungsbereich 386, Ludwigs-Maximilians-Universität München. Available under `http://www.stat.uni-muenchen.de/sfb386/publikation.html`.

Rue, H. (2000): Fast Sampling of Gaussian Markov Random Fields with Applications. Technical report. Available under `http://www.math.ntnu.no/preprint/statistics/2000/s1-2000.ps`.

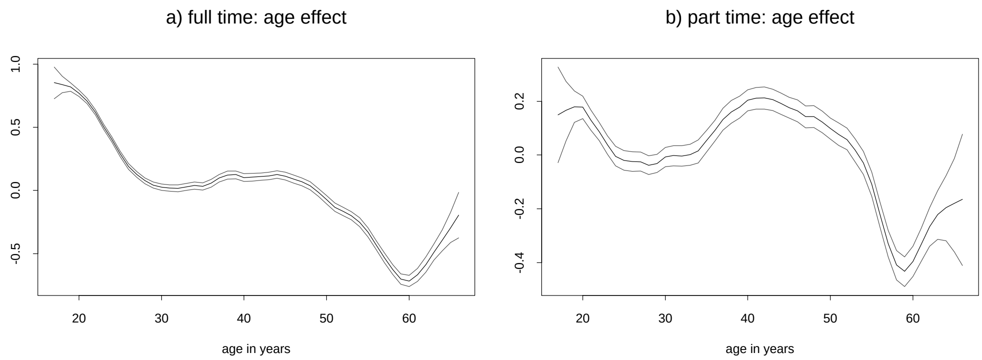


Figure 1: Estimated nonparametric effects of age. Shown is the posterior mean within 80 % credible regions

Variable	mean	Std.Dev.	10% quant.	median	90% quant.
german	0.0527878	0.00843951	0.0418975	0.0528639	0.0637414
foreign	-0.0527878	0.0084437	-0.0637414	-0.0528639	-0.0418975
unemployment assistance	-0.0733189	0.00687011	-0.0822028	-0.0735721	-0.0639811
unemployment benefit	0.0733189	0.0068735	0.063981	0.0735721	.0822028
no vocational training	-0.0342402	0.0104877	-0.0476529	-0.0346253	-0.0210024
vocational training	-.0195901	.0100236	-0.032336	-0.0193288	-.0069022
university	0.0538303	0.0183895	0.0295187	0.053873	0.077214
$P = 0$	-0.106041	0.0072038	-0.1157761	-0.1058226	-0.0968463
$P = 1$	-0.0361269	0.00862728	-0.047251	-0.0361909	-0.025397
$P = 2$	0.00232123	0.0111795	-0.0120529	0.00256533	0.0171483
$P \geq 3$	0.139847	0.0105637	0.125909	0.140088	0.153256

Table 2: estimates of constant parameters (full time)

Variable	mean	Std.Dev.	10% quant.	median	90% quant.
german	0.108616	0.0156238	0.0878499	0.10981	0.128572
foreign	-0.1086155	0.0156316	-0.1285725	-0.1098105	-0.0878499
unemployment assistance	-0.0839916	0.0112389	-0.0987005	-0.0836322	-0.0693947
unemployment benefit	0.0839916	0.0112445	0.0693947	0.0836322	0.0987005
no vocational training	-0.0989168	0.0150469	-0.118855	-0.0987753	-0.0798118
vocational training	-.0537657	.0143901	-.0723342	-.0532555	-.0347323
university	0.152682	0.0247098	0.120791	0.1523	0.185821
P=0	-.1592084	.0116431	-.1745401	-.1594094	-.1445379
P=1	-0.0506152	0.0135742	-0.0679307	-0.0507075	-0.0328039
P=2	0.00633397	0.0182364	-0.0172304	0.00684475	0.0298414
P=3	0.20349	0.0157018	0.183727	0.203319	0.223392

Table 3: estimates of constant parameters (part time)

	full time	part time
std. dev.	0.0246	0.103
minimum	-0.073	-0.274
10% quantile	-0.0287	-0.120
90% quantile	0.0291	0.134
maximum	0.128	0.501

Table 4: summary of the posterior means of the unstructured spatial effect

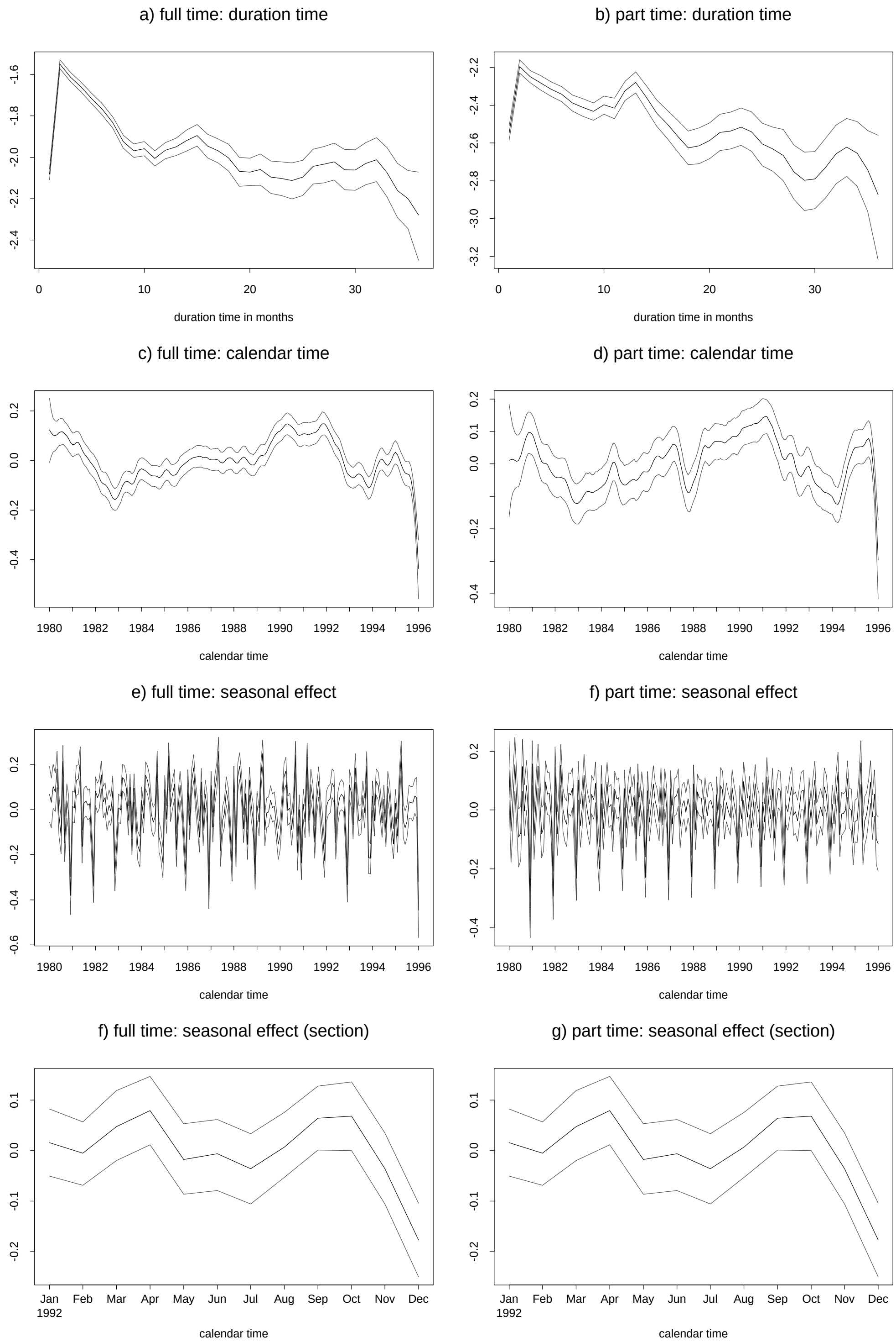


Figure 2: Estimated nonparametric effects of duration and calendar time. Shown is the posterior mean within 80 % credible regions

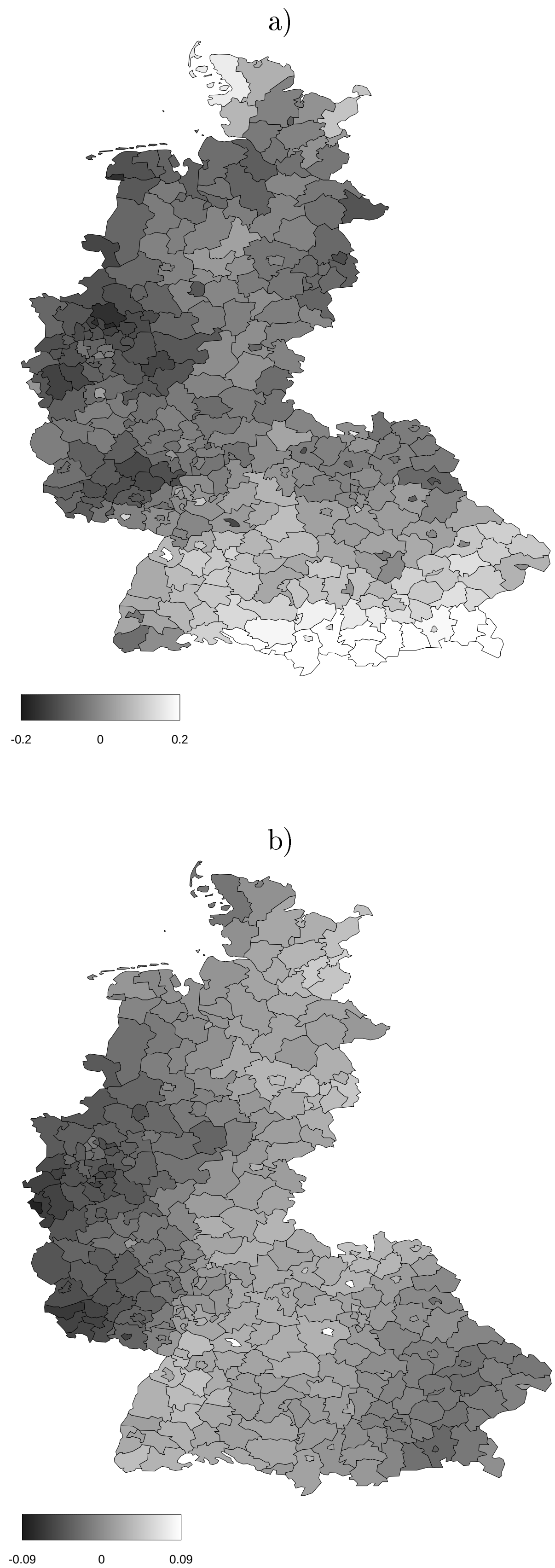
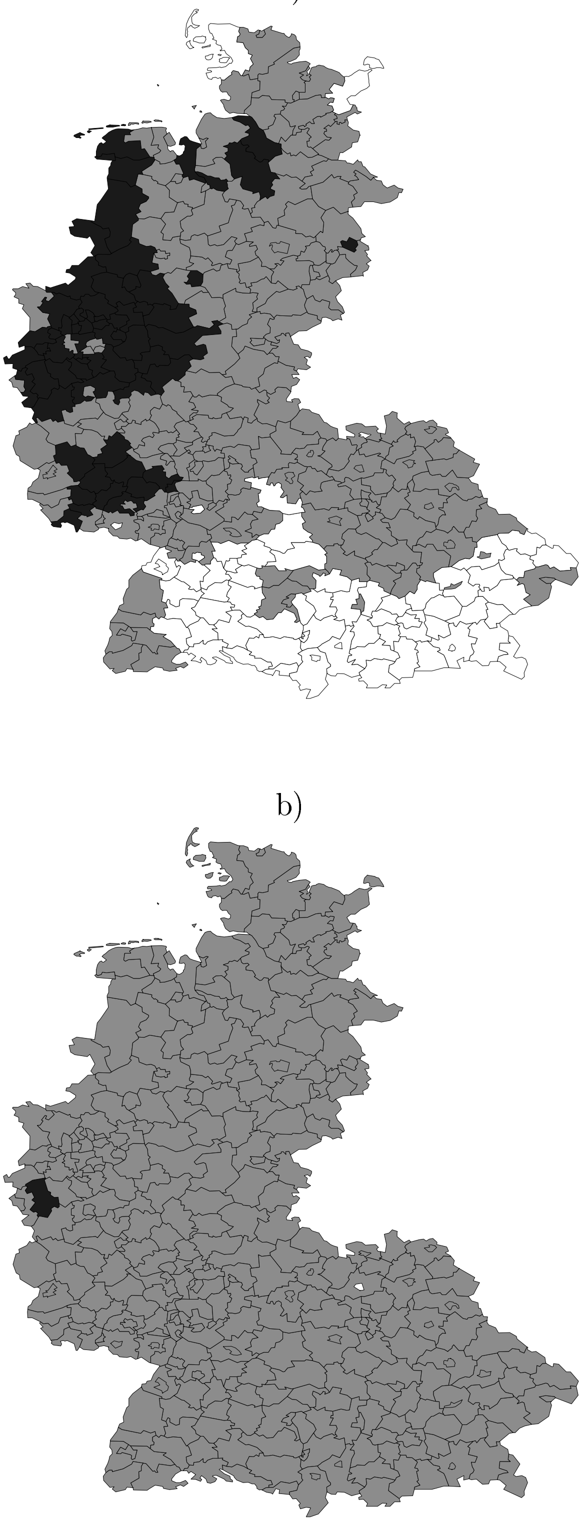


Figure 3: posterior mean of the structured spatial effects of the district specific effect

a)



b)

Figure 4: posterior probabilities of the district specific effect

Figure 5: Map of observation points

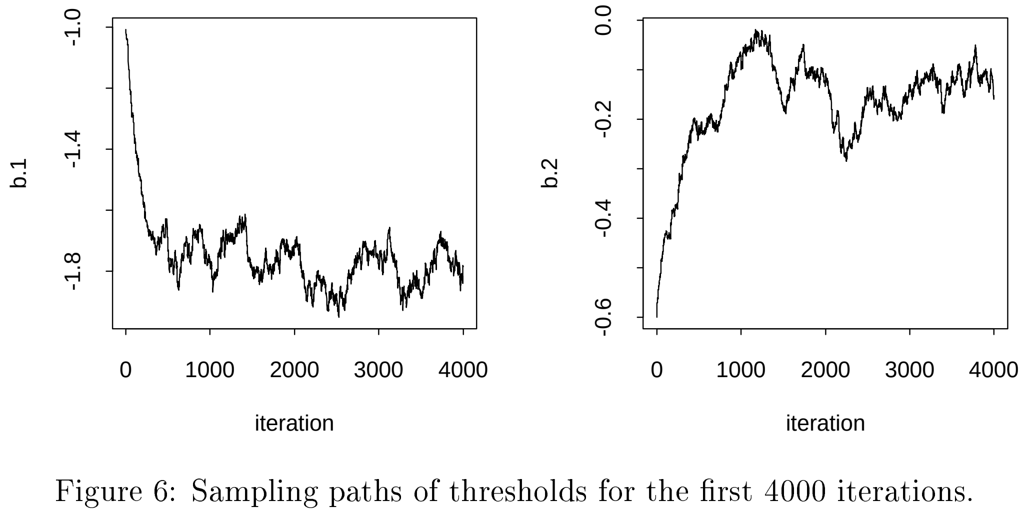


Figure 6: Sampling paths of thresholds for the first 4000 iterations.

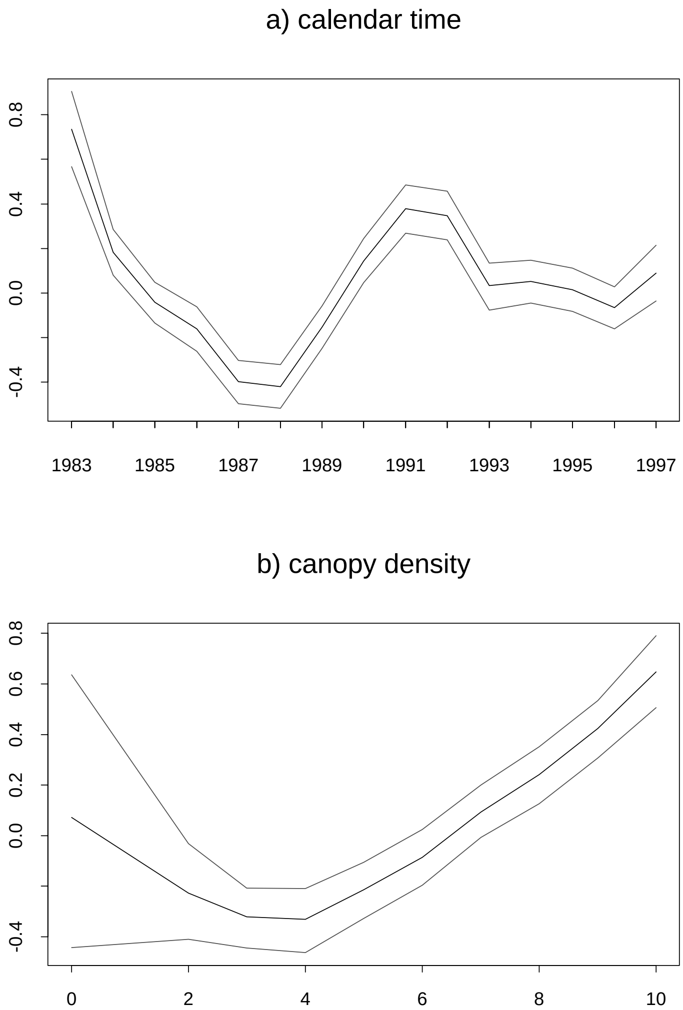


Figure 7: Estimated time trend and nonlinear effect of canopy density. Shown is the posterior mean within 80 % credible regions

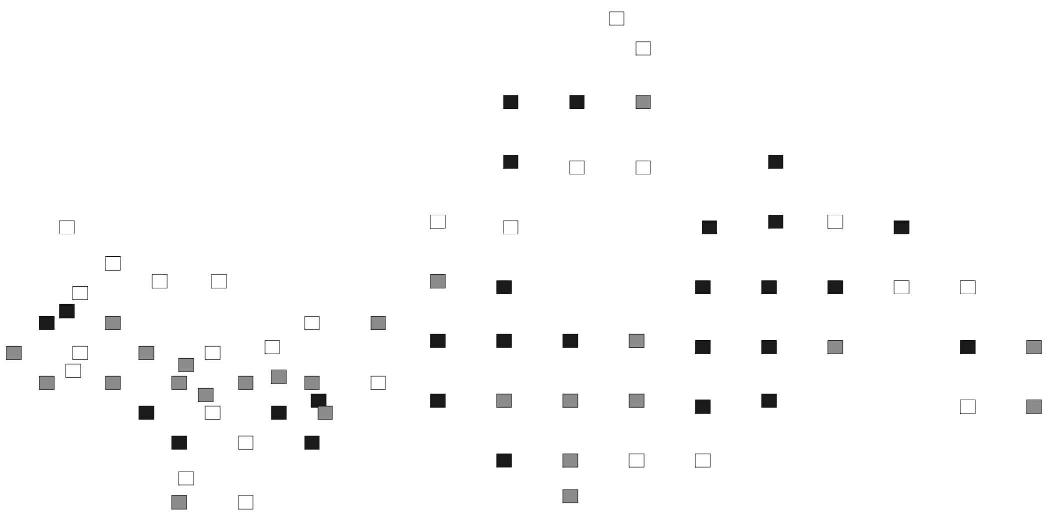


Figure 8: Estimated posterior probabilities of the spatial effect