



Hennерfeind, Brezger, Fahrmeir: Geoadditive survival models

Sonderforschungsbereich 386, Paper 333 (2003)

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

Projektpartner



Geoadditive survival models

Andrea Hennerfeind, Andreas Brezger, and Ludwig Fahrmeir

Department of Statistics, University of Munich

Abstract

Survival data often contain geographical or spatial information, such as the residence of individuals. We propose geoadditive survival models for analyzing spatial effects jointly with possibly nonlinear effects of other covariates. Within a unified Bayesian framework, our approach extends the classical Cox model to a more general multiplicative hazard rate model, augmenting the common linear predictor with a spatial component and nonparametric terms for nonlinear effects of time and metrical covariates. Markov random fields and penalized regression splines are used as basic building blocks. Inference is fully Bayesian and uses computationally efficient MCMC sampling schemes. Smoothing parameters are an integral part of the model and are estimated automatically. Performance is investigated through simulation studies. We apply our approach to data from a case study in London and Essex that aims to estimate the effect of area of residence and further covariates on waiting times to coronary artery bypass graft (CABG).

Key words: Bayesian hazard rate models, Markov random fields, penalized splines, MCMC, semi-parametric models, spatial survival data, CABG.

1 Introduction

In epidemiological, economic or social science applications, survival data often contain geographical or spatial information such as the district or postal code of the residence of individuals in the study. For example, Henderson, Shimakura and Gorst (2002) model spatial variation in survival of acute myeloid leukemia patients in northwest England, Banerjee, Wall and Carlin (2003) apply a spatial frailty model to infant mortality in Minnesota, and Li and Ryan (2002) analyze the effect of risk factors on the onset of childhood asthma with spatial data from the East Boston Asthma Study. In

a study on unemployment duration in Germany, Fahrmeir, Lang, Wolff and Bender (2003) investigate the impact of small area labor market regions and other covariates, such as calendar time, age and unemployment benefits. Because unemployment duration is given in months, they apply a geoadditive discrete-time probit model. With the same method, Adebayo and Fahrmeir (2003) model regional and socio-economic variation of childhood mortality in Nigeria. In Section 5 of this paper, we will apply our approach to data from the Appropriateness of Coronary Revascularisation (ACRE) study. Spatial survival data from this study have been recently analyzed within a discrete-time setting by Crook, Knorr-Held and Hemingway (2003).

In this paper, we propose flexible, continuous-time geoadditive hazard rate models. Within a unified framework, our approach extends the Cox model with respect to several aspects often needed in applications. We generalize the common linear predictor to an additive predictor, including unknown functional forms for the log-baseline hazard, time-varying effects and possibly nonlinear effects of metrical covariates, and a spatial component for geographical effects. In addition, uncorrelated frailty effects or nonlinear two-way interactions can be incorporated if appropriate. Modelling and inference is developed from a Bayesian perspective, using information from the full likelihood, rather than from a partial likelihood, in combination with priors for parameters and functions. Estimation of unknown functions of time and metrical covariates is based on penalized spline (P-spline) regression, introduced by Eilers and Marx (1996), Marx and Eilers (1998) for generalized additive models in a frequentist setting. We will use Bayesian versions (Lang and Brezger, 2003) as a basic building block. Basically, time is treated in the same way as a metrical covariate, but the degree and amount of smoothness may be different. For example, simple random walk priors for the log-baseline effect in a piecewise exponential model are P-splines of degree zero. The spatial component is modelled by Gaussian Markov random field priors, as common in disease mapping, or by two dimensional penalized tensor-product splines. A particular feature of our Bayesian approach is that all priors can be written in the same general form, see Section 2. This is also of advantage for a unified approach to posterior analysis with computationally efficient MCMC techniques. Smoothing parameters are an integral part of the model and can be estimated jointly with unknown functions and other parameters. Inferential procedures have been imple-

mented in C++ as part of *BayesX* (Brezger, Kneib and Lang, 2003).

Nonparametric Bayesian survival models have become quite popular in recent years, and some previous work deals with related, special cases of our approach. Ibrahim, Chen and Sinha (2001) provide a very good introduction and overview. Joint estimation of the baseline hazard and usual linear covariate effects in the Cox model has been considered by several authors. Gamerman (1991) proposes a Gaussian random walk model for the log–baseline hazard in the piecewise exponential model, and Sinha (1993) suggests a joint Gaussian smoothness prior. Arjas and Gasbarra (1994) introduce a first order autoregressive gamma model, and Cai, Hyndman and Wand (2002) use a mixed model representation of linear regression splines to estimate the baseline hazard. Time-varying effects have been treated within a state space framework by Gamerman (1991) for the piecewise exponential model. In all these approaches, however, main effects of metrical covariates are assumed to be of the usual linear form, and no spatial component is present. The spatial survival models of Li and Ryan (2002) and Henderson, Shimakura and Gorst (2002) add a spatial component in form of a stationary Gaussian field, common in geostatistics, to the linear predictor of the Cox model. Inference is based on partial likelihood by the first authors, while the second author plug in the Breslow estimator for the baseline hazard in MCMC updating steps. Carlin and Banerjee (2002) develop Bayesian spatio-temporal survival models, modelling baseline hazard functions nonparametrically through a beta mixture approach and assuming Markov random field priors for spatial effects. Nonparametric terms for nonlinear functions of time or other covariates are not included in these approaches.

Within the more conventional framework of Gaussian regression, geoadditive models have been suggested by Kamman and Wand (2003), who introduced the term "geoadditive". Inference is based on linear regression splines and stationary Gaussian fields, using a mixed model representation. Fully Bayesian modelling and inference for (generalized) geoadditive models has been developed in Fahrmeir and Lang (2001) and Brezger and Lang (2003). Because discrete-time survival models can be rewritten as a sequence of binary regression models, this approach can be adapted to discrete-time geoadditive survival models as in Crook et al. (2003) and Adebayo and Fahrmeir (2003).

The rest of the paper is organized as follows. In Section 2 we describe models, likelihood, and priors for unknown functions and parameters. Inference is outlined in Section 3. Performance is studied in Section 4 through a simulation study. Applications in Section 5 illustrate the method. The concluding section contains some proposals for future research.

2 Models, likelihood and priors

Consider survival data in usual form, i. e. , it is assumed that each individual i in the study has a lifetime T_i , and a censoring time C_i that are independent random variables. The observed lifetime is then $t_i = \min(T_i, C_i)$, and δ_i denotes the censoring indicator. The data are then given by

$$(t_i, \delta_i; v_i), \quad i = 1, \dots, n \quad (1)$$

where v is the vector of covariates. Covariates may also be time-dependent, but we restrict discussion to time-constant covariates for simplicity.

In Cox's proportional model the hazard rate for individual i is assumed as the product

$$\lambda_i(t; v_i) = \lambda_0(t) \exp(\gamma_1 v_{i1} + \dots + \gamma_r v_{ir}) = \lambda_0(t) \exp(v'_i \gamma). \quad (2)$$

The baseline hazard rate is unspecified, and, through the exponential link function, the covariates $v = (v_1, \dots, v_r)$ act multiplicatively on the hazard rate. As pointed out in the introduction, in a number of applications there is a need for extending this basic model with respect to several aspects. We propose novel nonparametric Bayesian survival models that can deal with these issues in a flexible and unified framework. Reparametrizing the baseline hazard rate through $\exp\{f_0(t)\}$, $f_0(t) = \log\{\lambda_0(t)\}$ and partitioning the vector of covariates into groups of covariates x, z, s and v , we extend model (2) to the nonparametric multiplicative model

$$\lambda_i(t) := \lambda_i(t; x_i, z_i, s_i, v_i) = \exp\{\eta_i(t)\} \quad (3)$$

with predictor

$$\eta_i(t) = f_0(t) + \sum_{j=1}^p f_j(t) z_{ij} + \sum_{j=p+1}^{p+q} f_j(x_{i,j-p}) + f_{spat}(s_i) + v'_i \gamma + b_{gi}. \quad (4)$$

Here $f_0(t) = \log\{\lambda_0(t)\}$ is the log-baseline effect, $f_j(t)$ are time-varying effects of covariates z_j , $f_j(x)$ is the nonlinear effect of a metrical covariate x , $f_{spat}(s)$ is the (structured) effect of a spatial

covariate s , with $s_i = s$ if unit i is from area s , $s = 1, \dots, S$, γ is the vector of usual linear fixed effects, and b_g is a unit- or group-specific frailty or random effect, with $b_{gi} = b_g$ if unit i is in group g , $g = 1, \dots, G$. For $G = n$, we obtain individual-specific frailties, for $G < n$, b_g might be the effect of centre g in a multicentre study or the unstructured (random) spatial effect of an area (i. e. $b_g = b_s$), for example. As an extension, random slopes could be introduced in (4), but we omit this here. Several other extensions of the model, such as choice of other link functions, inclusion of interactions and competing risks, are possible. We discuss this in the concluding section.

Under the usual assumption about noninformative censoring, the likelihood is given by

$$\begin{aligned} L &= \prod_{i=1}^n \lambda_i(t)^{\delta_i} \cdot \exp\left(-\int_0^{t_i} \lambda_i(u) du\right) \\ &= \prod_{i=1}^n \lambda_i(t)^{\delta_i} \cdot S_i(t), \end{aligned} \quad (5)$$

inserting (3) and (4).

The Bayesian model formulation is completed by assumptions about priors for parameters and functions. For fixed effect parameters γ we assume diffuse priors $p(\gamma) \propto \text{const}$. A weakly informative normal prior would be another choice. Random effects are assumed to be i.i.d. Gaussian, $b_g \sim N(0, \tau_b^2)$.

For unknown functions f_j , we assume Bayesian P-spline priors as in Lang and Brezger (2002). Random walk priors, which have been suggested in Fahrmeir and Lang (2001) and may be used as smoothness priors for the baseline effect and dynamic effects in a piecewise exponential model, appear as a special case. The basic idea of P-spline regression (Eilers and Marx, 1996) is to approximate a function $f_j(x)$ as a linear combination of B-spline basis functions B_m , i.e.

$$f_j(x) = \sum_{m=1}^{M_j} \beta_{jm} B_m(x).$$

The basis functions B_m are B-splines of degree l defined over a grid of equally spaced knots $x_{min} = \xi_0 < \xi_1 < \dots < \xi_s = x_{max}$, $M_j = l + s$. The number of knots is moderate, but not too small, to maintain flexibility, but smoothness of the function is encouraged by difference penalties for neighbouring coefficients in the sequence $\beta_j = (\beta_{j1}, \dots, \beta_{jM_j})'$. The Bayesian analogue are first or second order random walk smoothness priors

$$\beta_{jm} = \beta_{j,m-1} + u_{jm} \quad \text{or} \quad \beta_{jm} = 2\beta_{j,m-1} - \beta_{j,m-2} + u_{jm} \quad (6)$$

with i.i.d. Gaussian errors $u_{jm} \sim N(0, \tau_j^2)$ and diffuse priors $p(\beta_{j1}) \propto const$, or $p(\beta_{j1})$ and $p(\beta_{j2}) \propto const$, for initial values. A first order random walk penalizes abrupt jumps $\beta_{jm} - \beta_{j,m-1}$, and a second order random walk penalizes deviations from a linear trend. The amount of smoothness or penalization is controlled by the variance τ_j^2 , which acts as a smoothness parameter.

The joint prior of the regression parameters β_j is Gaussian and can be easily computed as a product of conditional densities defined by (6) as

$$\beta_j | \tau_j^2 \propto \exp\left(-\frac{1}{2\tau_j^2}\beta_j' K_j \beta_j\right).$$

The penalty matrix K_j is of the form $K_j = D'D$, where D is a first or second order difference matrix. For second order random walks, for example, K_j is given by

$$K_j = \begin{pmatrix} 1 & -2 & 1 & & & \\ -2 & 5 & -4 & 1 & & \\ 1 & -4 & 6 & -4 & 1 & \\ & 1 & -4 & 6 & -4 & 1 \\ & & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & 1 & -4 & 6 & -4 & 1 \\ & & & & 1 & -4 & 6 & -4 & 1 \\ & & & & & 1 & -4 & 5 & -2 \\ & & & & & & 1 & -2 & 1 \end{pmatrix}$$

with zero elements outside the second off-diagonals.

The band structure of K_j is very useful for computationally efficient MCMC updating schemes. A common choice for approximating smooth curves are quadratic or cubic B-splines. Computationally, linear splines are simpler. The simplest choice are B-splines of degree zero, i. e. $B_m(x) \equiv 1$ over the m -th interval, and $B_m(x) \equiv 0$ elsewhere. Then the effect is approximated by a piecewise constant function, and the function values follow a random walk model as in Fahrmeir and Lang (2001). This special choice, with time t as covariate, is the easiest way to smooth the baseline in the piecewise exponential model; moreover the integral in the likelihood (5) reduces to a sum, see the next section. With P-splines of higher degree, however, estimation of smooth baseline effects is improved in terms of MSE's, see Section 4.

For the structured spatial effect $f_{spat}(s)$ we assume either Markov random field priors or two dimensional tensor product P-spline priors. Considering small area data with sparse data for at least some of the areas, fixed area-specific effects would not lead to reliable estimations. Therefore we fit a smoothed spatial effect by using a technique that borrows strength from neighbouring areas, i. e. we assume that neighbouring areas are more similar than arbitrary areas and therefore the spatial effect varies smoothly.

In the case of MRF priors we define areas as neighbours if they share a common boundary and assume that the effect of an area s is normally distributed with the mean of the effects of neighbouring areas as expectation and a variance that is inverse proportional to the number of neighbours of area s , i. e.

$$f_{spat}(s) := \beta_s^{spat} = \frac{1}{N_s} \sum_{j \in \delta_s} \beta_j^{spat} + u_s, \quad u_s \sim N\left(0, \frac{\tau_s^2}{N_s}\right)$$

where N_s is the number of neighbours of area s and $j \in \delta_s$ denotes that area j is a neighbour of area s .

Considering the x- and y-coordinates of the geographical center of each area, the spatial effect can be seen as an interaction between two metrical covariates x_s and y_s . Therefore our second approach is based on two dimensional P-splines, a rather parsimonious, but flexible method for modelling interactions between metrical covariates described in Lang and Brezger (2003) for Gaussian regression. The assumption is that the unknown structured spatial effect $f_{spat}(s)$ can be approximated by the tensor product of one dimensional B-splines, i. e.

$$f_{spat}(s) = f_{spat}(x_s, y_s) = \sum_{m_1=1}^{M_s} \sum_{m_2=1}^{M_s} \beta_{s,m_1 m_2} B_{s,m_1}(x_s) B_{s,m_2}(y_s).$$

As before with one dimensional P-splines the B-splines of degree l are defined over a grid of a moderate, but not too small number of equally spaced knots. Smoothness is encouraged by a two dimensional first order random walk smoothness prior (with smoothing parameter τ_s^2) for $\beta_s = (\beta_{s,11}, \dots, \beta_{s,M_s M_s})'$. More details can be found in Lang and Brezger (2003).

Again, in both approaches the amount of smoothness is controlled by a smoothing parameter τ_s^2 that is estimated jointly with the unknown parameters β_s .

When applying our model to real data we do not know how much of the spatial variation is structured and how much is unstructured. Therefore we fit an additional (unstructured) area-

specific random effect and let the data decide.

Variances τ_j^2 follow inverse Gamma priors $IG(a_j; b_j)$. The hyperparameters a_j, b_j are chosen such that this prior is weakly informative. We routinely use $a_j = b_j = 0.001$ as a standard choice. For moderate to large data sets, results are rather insensitive to the choice of a_j and b_j . For smaller data sets, a sensitivity analysis is useful. We also assume an inverse Gamma prior for the variance τ_b^2 of the normal random effects b_g and the variance τ_s^2 of the spatial effect.

The Bayesian model specification is completed by assuming that all priors for parameters are conditionally independent, and that all priors are mutually independent.

3 Markov chain Monte Carlo inference

In what follows, let $\beta = (\beta_0, \dots, \beta_p, \beta_{p+1}, \dots, \beta_{p+q})'$ denote the vector of all B-spline regression coefficients, $\beta_s = (\beta_1^{spat}, \dots, \beta_S^{spat})'$ the vector of spatial effects in the case of a MRF-prior or $\beta_s = (\beta_{s,11}, \dots, \beta_{s,M_s M_s})'$ the coefficient vector of the two dimensional P-spline, respectively, γ the vector of fixed effects, b the vector of random effects, and $\tau^2 = (\tau_0^2, \dots, \tau_{p+q}^2, \tau_s^2, \tau_b^2)$ the vector of all variance components.

Full Bayesian inference is based on the entire posterior distribution

$$p(\beta, \beta_s, \gamma, b, \tau^2 \mid data) \propto L(\beta, \beta_s, \gamma, b, \tau^2) p(\beta, \beta_s, \gamma, b, \tau^2).$$

Due to the (conditional) independence assumptions, the joint prior factorizes into

$$p(\beta, \beta_s, \gamma, b, \tau^2) = \left\{ \prod_{j=0}^{p+q} p(\beta_j \mid \tau_j^2) p(\tau_j^2) \right\} p(\beta_s \mid \tau_s^2) p(\tau_s^2) \left\{ \prod_{g=1}^G p(b_g \mid \tau_b^2) \right\} p(\tau_b^2) p(\gamma),$$

where the last factor can be omitted for diffuse fixed effect priors.

The likelihood $L(\beta, \beta_s, \gamma, b, \tau^2)$ is given by inserting (3),(4) into (5), but the integral requires integration over all terms depending on survival time t , i. e. terms of the form

$$I_i = \int_0^{t_i} \exp \left(f_0(u) + \sum_{j=1}^p f_j(u) z_{ij} \right) du,$$

where $f_j(t) = \sum \beta_{jm} B_m(t)$. Apart from B-splines $B_m(t)$ of degree zero, i. e. random walk models, and linear B-splines, these integrals are not available in closed form. The first case leads to the

piecewise exponential model: The time axis is divided into a grid

$$0 = \xi_0 < \xi_1 < \dots < \xi_{t-1} < \xi_t < \dots < \xi_s = t_{max},$$

and $f_j(t)$ is assumed to be a piecewise constant function, i. e.

$$f_j(t) = \beta_{jt}$$

in time interval $(\xi_{t-1}, \xi_t]$, $t = 1, \dots, s$. In this case, the integral reduces to a sum, and, after some simple calculations, the likelihood can rewritten in the form of a Poisson–likelihood, with the predictor η_{it} containing an additional offset term, see Fahrmeir and Tutz (2001, Section 9.1) or Ibrahim, Chen and Sinha (2001, Section 3.1) for details.

For linear B–splines, the integrals can still be solved analytically, but expressions are rather messy and the computational effort is quite high, see Cai et al. (2002, Appendix). Following their suggestion, we use simple numerical integration in form of the trapezoidal rule for linear B–splines as well as for the commonly used cubic B–splines, where analytical integration is not possible anyway. Full Bayesian inference via MCMC simulation is based on updating full conditionals of single parameters or blocks of parameters, given the rest of the data.

For updating the parameter vectors β_j , $j = p + 1, \dots, p + q$, which correspond to the time–independent functions $f_j(x)$, as well as spatial effects β_s (the parameter vector β_s of the two dimensional P–spline, respectively), fixed effects γ and random effects b , we use a slightly modified version of an MH–algorithm based on iteratively weighted least squares (IWLS) proposals, developed for fixed and random effects by Gamerman (1997) and adapted to generalized additive mixed models in Brezger and Lang (2003). More precisely, the goal is to approximate the posterior by a Gaussian distribution, obtained by accomplishing *one* IWLS step in every iteration of the sampler. Then, random samples have to be drawn from a high dimensional multivariate Gaussian distribution with precision matrix and mean

$$P_j = X'_j W(\beta_j^c) X_j + \frac{1}{\tau_j^2} K_j, \quad m_j = P_j^{-1} X'_j W(\beta_j^c) (\tilde{y} - \tilde{\eta}).$$

Here, $\tilde{\eta}_i = \eta_i(t_i) - f_j(x_{i,k-p})$, $W(\beta_j^c) = diag(w_1, \dots, w_n)$ is the weight matrix for IWLS with weights

$$w_i = \exp \left(\sum_{k=p}^{p+q} f_k(x_{i,k-p}) + f_{spat}(s_i) + v'_i \gamma + b_{g_i} \right) \cdot I_i$$

obtained from the current state β_j^c . The working observations \tilde{y}_i are given by

$$\tilde{y}_i = \eta_i(t_i) + \frac{\delta_i}{w_i} - 1.$$

Random numbers from the high dimensional proposal distributions can be efficiently drawn by using matrix operations for sparse matrices.

Suppose we want to update β_j , with current value β_j^c of the chain. Then a new value β_j^p is proposed by drawing a random vector from a (high-dimensional) multivariate Gaussian proposal distribution $q(\beta_j^c, \beta_j^p)$, which is obtained from a quadratic approximation of the log-likelihood by a second order Taylor expansion with respect to β_j^c , in analogy to IWLS iterations in generalized linear models. The proposed vector β_j^p is accepted as the new state of the chain with probability

$$\alpha(\beta_j^c, \beta_j^p) = \min \left(1, \frac{p(\beta_j^p | \cdot) q(\beta_j^p, \beta_j^c)}{p(\beta_j^c | \cdot) q(\beta_j^c, \beta_j^p)} \right)$$

where $p(\beta_j | \cdot)$ is the full conditional for β_j (i. e. the conditional distribution of β_j given all other parameters and the data).

For a fast implementation, we use the fact that the precision matrices of the Gaussian proposal distributions are banded, so that Cholesky decompositions can be performed efficiently.

For the parameters β_0, \dots, β_p corresponding to the functions $f_0(t), \dots, f_p(t)$ depending on time t , the IWLS-MH algorithm requires considerably more computational effort, because the integrals in the log-likelihood as well as first and second derivatives are involved now. Therefore, we adopt a computationally faster MH-algorithm based on conditional prior proposals, although IWLS-MH has better mixing properties. This algorithm was first developed by Knorr-Held (1999) for state space models and extended for generalized additive mixed models in Fahrmeir and Lang (2001). It requires only evaluation of the log-likelihood, not of derivatives. However, draws are not performed for the entire vector β_j , but iteratively for blocks of subvectors, see Fahrmeir and Lang (2001) for details.

The full conditionals for the variance parameters τ_j^2 are inverse gamma with parameters

$$a'_j = a_j + \frac{1}{2} \text{rank}(K_j) \quad \text{and} \quad b'_j = b_j + \frac{1}{2} \beta'_j K_j \beta_j$$

and updating can be done by simple Gibbs steps, drawing random numbers directly from the inverse gamma densities. In complete analogy, the full conditional for a variance component τ_s^2 of the spatial effect and τ_b^2 of a random intercept or slope is again an inverse gamma distribution, and updating is straightforward.

For model comparison we suggest to use the Deviance Information Criterion (DIC) developed in Spiegelhalter et al. (2002). It is given as

$$DIC = D(\bar{\theta}) + 2p_D = \overline{D(\theta)} + p_D,$$

where θ is the vector of parameters, $D(\bar{\theta})$ is the deviance of the model evaluated at the posterior mean estimate $\bar{\theta}$, $\overline{D(\theta)}$ is the posterior mean of the deviance and $p_D = \overline{D(\theta)} - D(\bar{\theta})$ is the effective number of parameters. Since it is at least unclear, how the saturated model should be defined in the case of survival data, when the baseline hazard and other nonparametric functions are parameters of interest, we use the unstandardized deviance $D(\theta) = -2 \cdot \log\text{Likelihood}$ instead of the saturated deviance.

4 Simulation study

We investigate performance through a simulation study. Life times T_i , $i = 1, \dots, 1236$, were generated according to the hazard model

$$\begin{aligned} \lambda_i(t) &= \lambda_0(t) \exp(f_1(x_i) + f_{spat}(s_i) + \gamma v_i) \\ &= \exp(\log(3t^2) + \sin(x_i) + \sin(x_{s_i} \cdot y_{s_i}) - 0.3v_i). \end{aligned} \quad (8)$$

In this model, the baseline hazard rate $\lambda_0(t)$ is set to $3t^2$, which is a Weibull hazard rate, so that $f_0(t) = \log(3t^2)$. The covariate v is binary, with the v_i 's randomly drawn from a Bernoulli $B(1; 0.5)$ distribution, and the covariate x is continuous, with the x_i 's randomly drawn from a uniform $U[-3, 3]$ distribution. The spatial covariate s_i denotes one of the $s = 1, \dots, S = 309$ counties of the former Federal Republic of Germany and x_{s_i} and y_{s_i} are the centered coordinates of the geographic center of county s_i . We simulated four observations per county. Censoring variables C_i , $i = 1, \dots, 1236$, were generated as i. i. d. draws from a uniform $U[0, 5]$ distribution, resulting in a proportion of 15–20 percent of censored observations.

Keeping the predictor fixed, 100 replications $\{T_i^{(r)}, C_i^{(r)}, i = 1, \dots, 1236\}$ resp. $\{(t_i^{(r)}, \delta_i^{(r)}), i = 1, \dots, 1236\}$, $r = 1, \dots, 100$ of censored survival times were generated.

The log–baseline hazard $f_0(t)$ was modelled by second order random walk priors, corresponding to a piecewise exponential model (with grid length $\Delta = 0.1$), and – alternatively – as a cubic P–spline, with 20 knots. A cubic P–spline prior with 20 knots was chosen for $f_1(x) = \sin(x)$. The spatial effect was modelled as a MRF and alternatively as a two dimensional cubic P–spline with 12x12 knots. Hyperparameters of inverse gamma priors for variance components were set to $a = 0.001$, $b = 0.001$, the standard choice.

For each replication $r = 1, \dots, 100$, we computed the mean square errors

$$MSE_r(f_0) = \frac{1}{1236} \sum_{i=1}^{1236} (\hat{f}_0^{(r)}(t_i^{(r)}) - f_0(t_i^{(r)}))^2,$$

for the log–baseline hazard $f_0(t)$,

$$MSE_r(f_1) = \frac{1}{1236} \sum_{i=1}^{1236} (\hat{f}_1^{(r)}(x_i) - f_1(x_i))^2$$

for $f_1(x) = \sin(x)$, and

$$MSE_r(f_{spat}) = \frac{1}{1236} \sum_{i=1}^{1236} (\hat{f}_{spat}^{(r)}(s_i) - f_{spat}(s_i))^2$$

for the spatial effect $f_{spat}(s) = \sin(x_c \cdot y_c)$, where $\hat{f}_k^{(r)}$, $k = 0, 1, spat$, are posterior mean estimates for simulation run r .

The $MSE(\gamma)$ was computed in the usual way.

Table 2 summarizes the results, displaying $\overline{MSE} = \frac{(\sum_{r=1}^{100} MSE_r)}{100}$ as well as $\min_r MSE_r$ and $\max_r MSE_r$ in each cell.

As was to be expected, the P–spline model has smaller MSE ’s for f_0 when compared to the piecewise exponential model. Interestingly, the MSE ’s for $\gamma = -0.3$, $f_1(x)$ and $f_{spat}(s)$ are more or less unaffected by the choice of the smoothness prior for the log–baseline $f_0(t)$. Estimated functions of replication r , with r chosen such that MSE_r is the median of MSE_1, \dots, MSE_{100} , for $f_0(t)$, $f_1(x)$ and $f_{spat}(s)$ are displayed in Figures 7-9 in the appendix.

	MRF	geospline
piecewise exponential model	$\overline{MSE}(f_0) = 0.154$	$\overline{MSE}(f_0) = 0.155$
	$minMSE(f_0) = 0.049$	$minMSE(f_0) = 0.044$
	$maxMSE(f_0) = 0.497$	$maxMSE(f_0) = 0.496$
	$\overline{MSE}(f_1) = 0.0068$	$\overline{MSE}(f_1) = 0.0061$
	$minMSE(f_1) = 0.0006$	$minMSE(f_1) = 0.0006$
	$maxMSE(f_1) = 0.0193$	$maxMSE(f_1) = 0.0182$
	$\overline{MSE}(f_{spat}) = 0.042$	$\overline{MSE}(f_{spat}) = 0.022$
	$minMSE(f_{spat}) = 0.028$	$minMSE(f_{spat}) = 0.010$
	$maxMSE(f_{spat}) = 0.068$	$maxMSE(f_{spat}) = 0.039$
	$\overline{MSE}(\gamma) = 0.0045$	$\overline{MSE}(\gamma) = 0.0038$
P-spline model	$minMSE(\gamma) \approx 0$	$minMSE(\gamma) \approx 0$
	$maxMSE(\gamma) = 0.0268$	$maxMSE(\gamma) = 0.0197$
	$\overline{MSE}(f_0) = 0.126$	$\overline{MSE}(f_0) = 0.127$
	$minMSE(f_0) = 0.033$	$minMSE(f_0) = 0.027$
	$maxMSE(f_0) = 0.450$	$maxMSE(f_0) = 0.453$
	$\overline{MSE}(f_1) = 0.0070$	$\overline{MSE}(f_1) = 0.0063$
	$minMSE(f_1) = 0.0009$	$minMSE(f_1) = 0.0006$
	$maxMSE(f_1) = 0.0209$	$maxMSE(f_1) = 0.0178$
	$\overline{MSE}(f_{spat}) = 0.043$	$\overline{MSE}(f_{spat}) = 0.022$
	$minMSE(f_{spat}) = 0.029$	$minMSE(f_{spat}) = 0.010$
	$maxMSE(f_{spat}) = 0.071$	$maxMSE(f_{spat}) = 0.039$
	$\overline{MSE}(\gamma) = 0.0046$	$\overline{MSE}(\gamma) = 0.0038$
	$minMSE(\gamma) \approx 0$	$minMSE(\gamma) \approx 0$
	$maxMSE(\gamma) = 0.0297$	$maxMSE(\gamma) = 0.0202$

Table 1: Summary of MSE's

5 Application: waiting times to CABG

We illustrate our methods by an application to data from a study in London and Essex that aims to analyze the effects of area of residence and further individual specific covariates on waiting times to coronary artery bypass graft (CABG). The data comprise observations for 3015 patients with definite coronary artery disease who were referred to one cardiothoracic unit from five contiguous health authorities. Waiting times from angiography to CABG are given in days. Covariates are, among others, sex, age (in years), numbers of diseased vessels (1, 2, 3), and the area of residence (one of 488 electoral wards).

The data were previously analyzed by Crook, Knorr–Held and Hemingway (2003) who classified waiting times in months and applied discrete–time survival methodology as described for examples in Fahrmeir and Tutz (2001, ch.9). They analyzed and compared a hierarchy of models, with model comparison based on the deviance information criterion (DIC), developed in Spiegelhalter et al. (2002). Here we apply continuous–time geoadditive survival models, with waiting times given in days as in the original data set, and predictors based on models 11 and 12 in Crook et al. (2003), which were the best in terms of DIC. Model 11 corresponds to a continuous–time model with hazard rate

$$\lambda(t) = \exp(f_0(t) + f_{\text{age}}(\text{age}) + f_s(\text{ward}) + b_{\text{ward}} + \gamma_1 \text{sex} + \gamma_2 \text{dv2} + \gamma_3 \text{dv3}), \quad (9)$$

where $f_0(t)$ is the log–baseline rate, $f_{\text{age}}(\text{age})$ is the nonlinear effect of age, $f_s(\text{ward})$ is the structured spatial effect modelled through a MRF prior, and b_{ward} is the unstructured spatial effect of ward, modelled through uncorrelated random effects. The remaining covariates are dummy–coded: sex=1 for female, and sex=0 for male, dv2=1 if the number of diseased vessels=2, dv2=0 else, and dv3=1 if the number of diseased vessels=3, dv3=0 else.

The (log–) baseline prior was assumed as a (log–) piecewise exponential model with grid length $\Delta=50$ days and, alternatively, as a cubic P–spline model with 20 knots. For f_{age} we assumed a cubic P–spline prior with 20 knots.

Model 12 is a modification of (9), where the fixed effects γ_2 and γ_3 of dv2 and dv3 are replaced by

time varying effects and the unstructured spatial effect of ward is removed:

$$\lambda(t) = \exp(f_0(t) + f_{\text{age}}(\text{age}) + f_s(\text{ward}) + \gamma_1 \text{sex} + f_1(t)\text{dv2} + f_2(t)\text{dv3}). \quad (10)$$

This model is a non-proportional hazard model and can be compared to the geoadditive proportional hazard rate model (9).

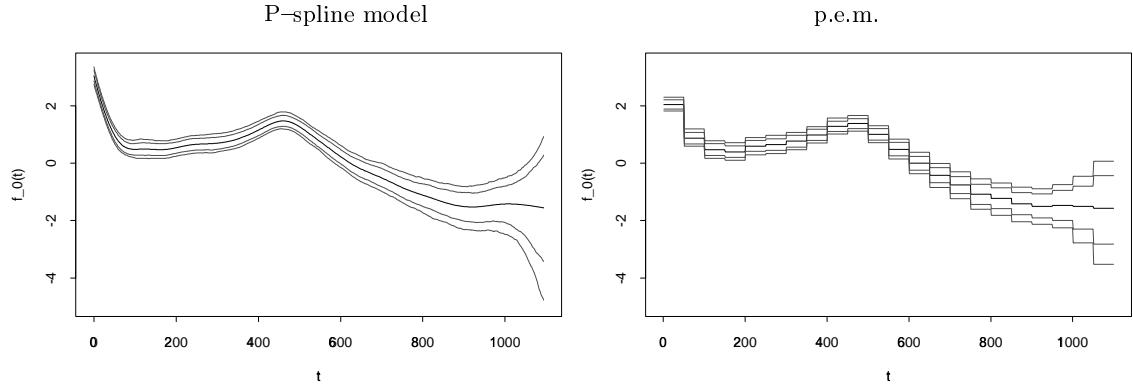


Figure 1: Posterior mean estimate for the (log-)baseline effect on time to CABG and 80% and 95% credible intervals

	P-spline model	p.e.m.
sex	-0.037 (0.082)	-0.040 (0.085)
dv2	1.496 (0.100)	1.500 (0.094)
dv3	1.815 (0.095)	1.817 (0.089)
const	-9.382 (0.213)	-9.356 (0.198)

Table 2: Posterior mean estimates and standard deviations for the fixed effects on time to CABG

Table 2 contains estimation results for the fixed effects in model (9). While the effect of sex is nonsignificant, the effects of two or three diseased vessels are clearly significant and show that waiting times are decreasing with increasing number of vessels. These results correspond to the findings of Crook et al. (2003). The baseline effects in Figure 1 show an initially high, but strongly decreasing chance of CABG immediately after diagnosis, followed by a slow increase between 150 - 450 days. Later, the chance of being operated decreases. The overall pattern is similar to the

results in Crook et al. (2003), obtained with a discrete-time model. However, with the P-spline prior we get a distinctly smoother curve. The effect of age (Figure 2) is almost constant between 40 and 80 years and does not have significant influence on the waiting time. Also, the estimates under a piecewise exponential and a cubic P-spline baseline prior are visually indistinguishable.

This is also true for the structured spatial effects shown in Figure 3. The maps give an

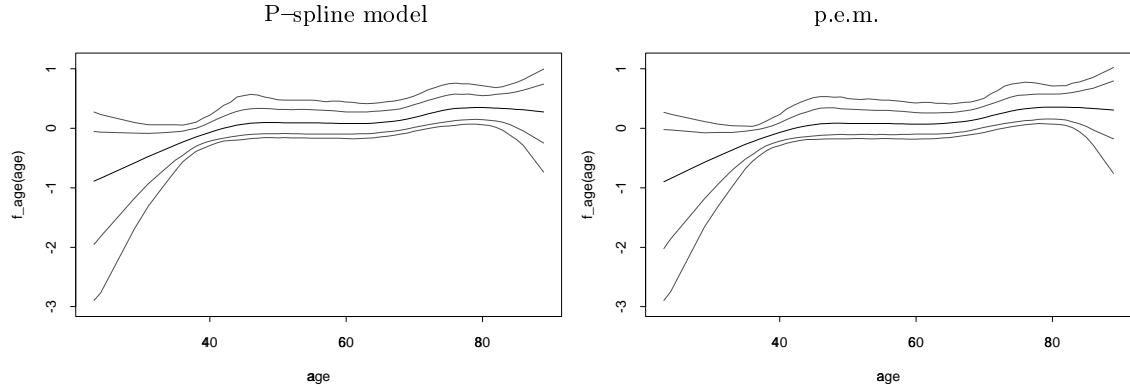


Figure 2: Posterior mean estimates of the effect of age on time to CABG and 80% and 95% credible intervals

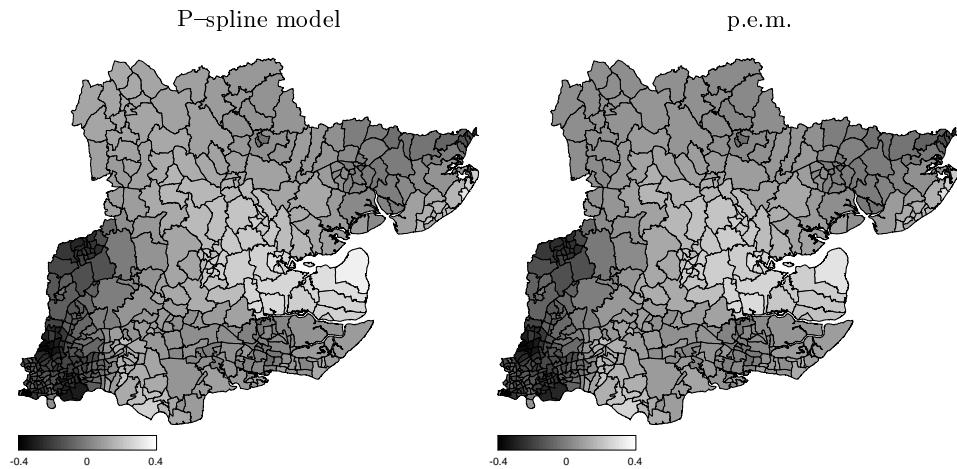


Figure 3: Posterior mean estimates of the structured spatial effect on time to CABG

impression of the spatially varying chance of CABG with light (dark) areas indicating an increased (decreased) effect. Areas with increased chances are Chelmsford and Malden in North Essex, while in areas around Harlow in North Essex and Walthamstow and Chingford in North East London chances are lower, that means patients have to wait longer for surgery. The maps in Figure 4 show

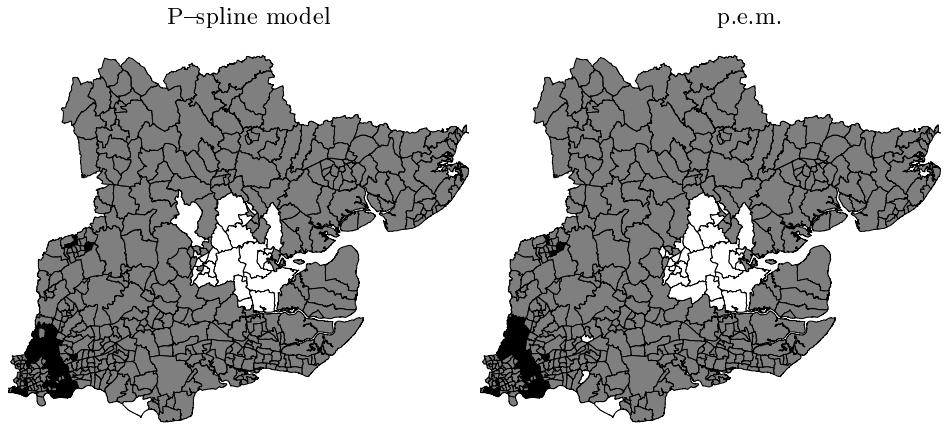


Figure 4: Posterior probabilities of the structured spatial effects, with white (black) areas indicating that at least 80 % of the sample estimates were positive (negative)

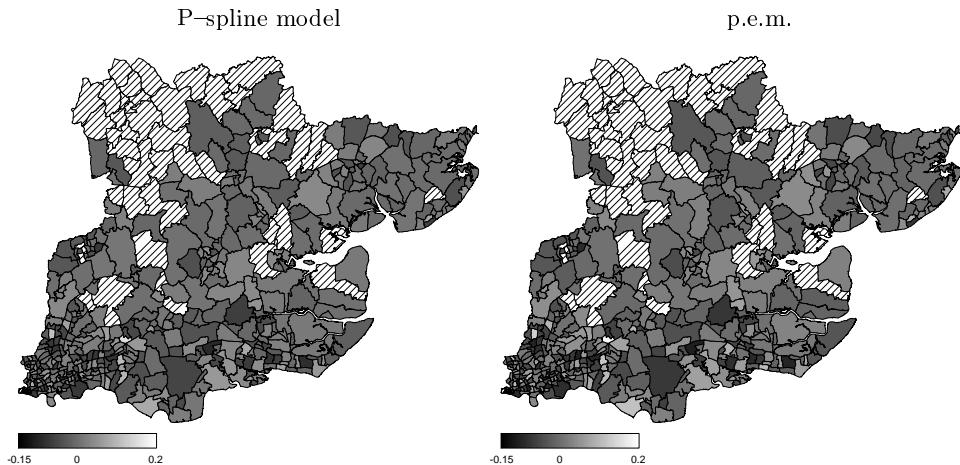


Figure 5: Posterior mean estimates of the random spatial effect on time to CABG

posterior probabilities of these spatial effects. White (black) areas indicate that at least 80 per cent of the sample estimates were positive (negative). Remaining grey areas are considered as 'non-significant'. The unstructured random effects in Figure 5 are much smaller than the structured spatial effects and cannot be reasonably interpreted. Therefore, unstructured spatial effects have been omitted in the final analysis of model (10).

Model (10) with time-varying effects $f_1(t)$ and $f_2(t)$ of dv2 and dv3 can be interpreted as a model with three separate baseline effects $f_0(t)$, $f_0(t) + f_1(t)$, $f_0(t) + f_2(t)$ for patients with one, two or three diseased vessels, respectively. The corresponding estimated curves are displayed in Figure 6 and indicate that the proportional hazards assumption is violated, because the baseline effect of

patients with three diseased vessels crosses the two other curves.

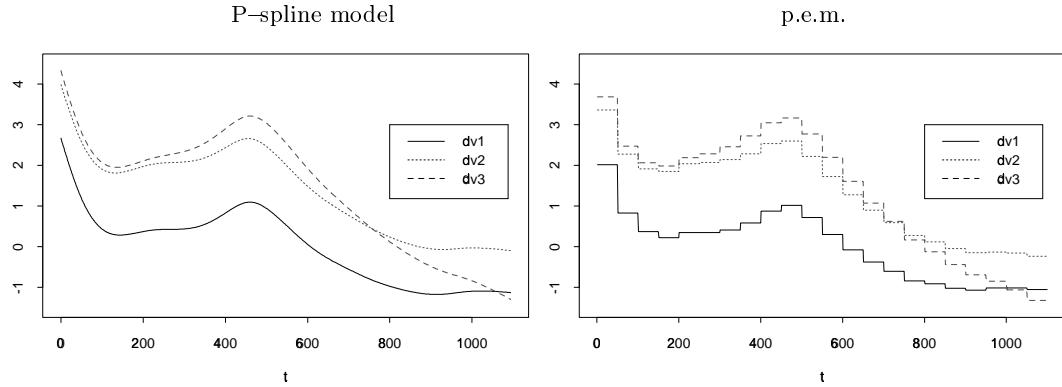


Figure 6: (log-)baseline effects on time to CABG: posterior mean estimates for 1 diseased vessel (dv1), 2 diseased vessels (dv2) and 3 diseased vessels (dv3)

6 Conclusion

Spatial extensions of statistical models for analyzing survival and, more general, event history data, will be of increasing relevance because spatial small-area information is often available. Assessment of spatial effects on hazard or survivor functions is not only of interest in its own but can be quite useful for detecting unobserved covariates, which carry spatial information.

In this work, we have developed a flexible class of nonparametric geoadditive survival models within a unified Bayesian framework for modelling and inference. Several extensions could be considered in future research. More general event history models and censoring mechanisms including spatial components can be embedded in the counting process framework. An important practical issue is the development of numerically efficient solutions for evaluating the likelihood in the presence of time-varying effects of covariates. A related point is to explore properties of criteria for model choice such as the DIC.

Acknowledgement

We thank Dr. H. Hemingway and A. Crook for providing the data for the CABG study, Stefan Lang for helpful discussions and computational assistance with *BayesX* and Leonhard Held for helpful comments on the CABG data. Financial support of the German Science Foundation DFG, Sonderforschungsbereich 386 "Statistische Analyse Diskreter Strukturen" is gratefully acknowledged.

References

- Adebayo, S. and Fahrmeir, L., 2003: Analyzing child mortality in Nigeria with geoadditive survival models. Discussion Paper 303, SFB 386, Ludwig-Maximilians-Universität München.
- Arjas, E. and Gasbarra, D., 1994: Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica* 4, 505–524.
- Banerjee, S., Wall, M. M., and Carlin, B. P., 2003: Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4, 123–142.
- Brezger, A., Kneib, T. and Lang, S., 2003: BayesX - Software for Bayesian Inference based on Markov Chain Monte Carlo simulation techniques, open domain software available from <http://www.stat.uni-muenchen.de/~lang/>.
- Brezger, A. and Lang, S., 2003: Generalized structured additive regression based on Bayesian P-splines. Discussion Paper 321, SFB 386.
- Cai, T., Hyndman, R., and Wand, M., 2002: Mixed model-based hazard estimation. *Journal of Computational and Graphical Statistics*, 11, 784–798.
- Carlin, B. P., and Banerjee, S., 2002: Hierarchical Multivariate CAR Models for Spatio-Temporally Correlated Data. *Bayesian Statistics 7*, eds. J.M. Bernardo et al., Oxford: Oxford University Press, to appear.

Crook, A., Knorr-Held, L. and Hemingway, H., 2003: Measuring spatial effects in time to event data: a case study using months from angiography to coronary artery bypass graft (CABG). *Statistics in Medicine*, to appear.

Eilers, P.H.C. and Marx, B.D., 1996: Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder). *Statistical Science*, 11 (2), 89-121.

Fahrmeir, L. and Lang, S., 2001: Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors. *Journal of the Royal Statistical Society C (Appl. Stat.)*, 50, 201-220.

Fahrmeir, L., Lang, S., Wolff, J. and Bender, S. (2003): Semiparametric Bayesian Time-Space Analysis of Unemployment Duration. *Journal of the German Statistical Society (Allgemeines Statistisches Archiv)*, to appear.

Fahrmeir, L. and Tutz, G., 2001: *Multivariate Statistical Modelling based on Generalized Linear Models*, Springer-Verlag, New York.

Gamerman, D., 1991: Dynamic Bayesian models for survival data. *Applied Statistics*, 40, 63–79.

Gamerman, D., 1997: Efficient Sampling from the Posterior Distribution in Generalized Linear Models. *Statistics and Computing*, 7, 57–68.

Henderson, R., Shimakura, S. and Gorst, D., 2002: Modeling Spatial Variation in Leukemia Survival Data. *Journal of the American Statistical Association*, to appear.

Ibrahim, J., Chen, M. H. and Sinha, D., 2001: *Bayesian Survival Analysis*. Springer Series in Statistics, New York.

Kamman, E. E. and Wand, M. P., 2003: Geoadditive models, *Applied Statistics*, in press.

Knorr-Held, L., 1999: Conditional Prior Proposals in Dynamic Models. *Scandinavian Journal of Statistics*, 26, 129–144.

Lang, S. and Brezger, A., 2003: Bayesian P-splines. *Journal of Computational and Graphical Statistics*, to appear.

Li, Y. and Ryan, L., 2002: Modeling Spatial Survival Data Using Semiparametric Frailty Models. *Biometrics*, 58, 287–297.

Marx, B. D. and Eilers, P., 1998: Direct Generalized Additive Modeling with Penalized Likelihood. *Computational Statistics and Data Analysis*, 28, 193–209.

Sinha, D., 1993: Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association*, 92, 1195–1212.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A., 2002: Bayesian measures of model complexity and fit (with discussion and rejoinder). *J. Roy. Statist. Soc., Ser. B*, 64, 583–639.

Appendix

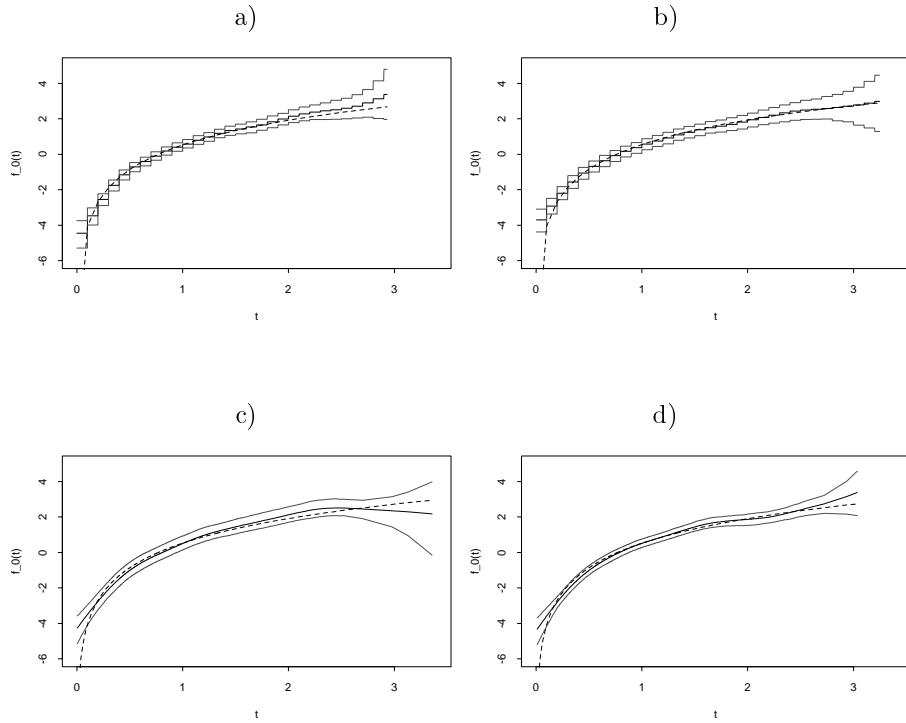


Figure 7: (log-)Baseline effects for the various model specifications; displayed are posterior mean estimates and 95 % credible intervals of run r , with r chosen such that MSE_r is the median of MSE_1, \dots, MSE_{100} (solid lines), and the true (log-)baseline effect (dashed line).

- a) p.e.m., MRF, $r=59$, $MSE=0.138$
- b) p.e.m., geospline, $r=4$, $MSE=0.140$
- c) P-spline model, MRF, $r=32$, $MSE=0.106$
- d) P-spline model, geospline, $r=24$, $MSE=0.112$

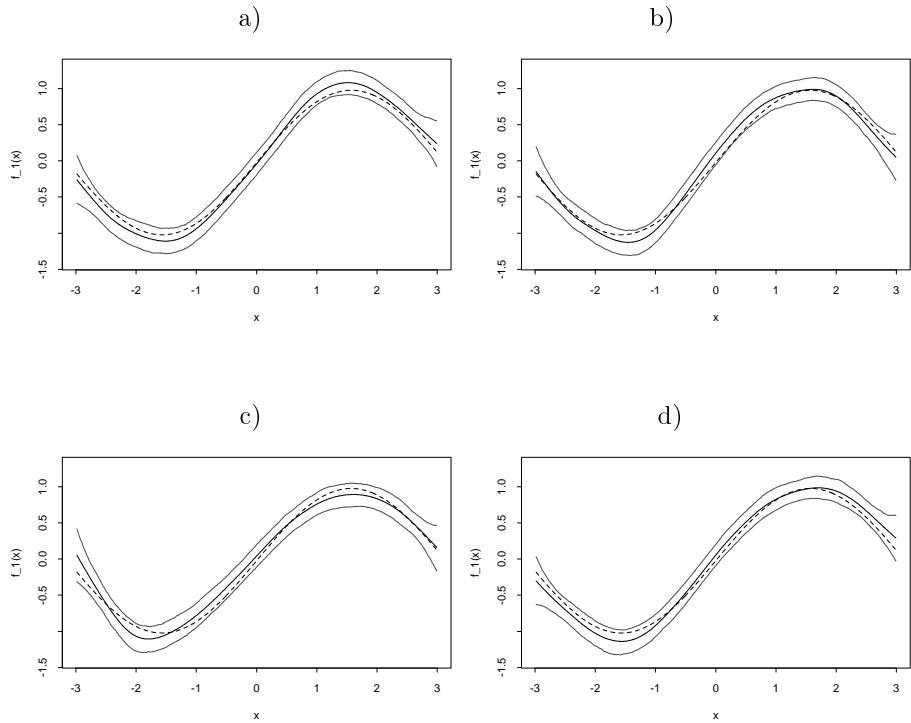


Figure 8: Nonparametric effects for the various model specifications; displayed are posterior mean estimates and 95% credible intervals of run r , with r chosen such that MSE_r is the median of MSE_1, \dots, MSE_{100} (solid lines), and the true function (dashed line).

- a) p.e.m., MRF, $r=9$, $MSE=0.0059$
- b) p.e.m., geospline, $r=70$, $MSE=0.0056$
- c) P-spline model, MRF, $r=26$, $MSE=0.0061$
- d) P-spline model, geospline, $r=74$, $MSE=0.0057$

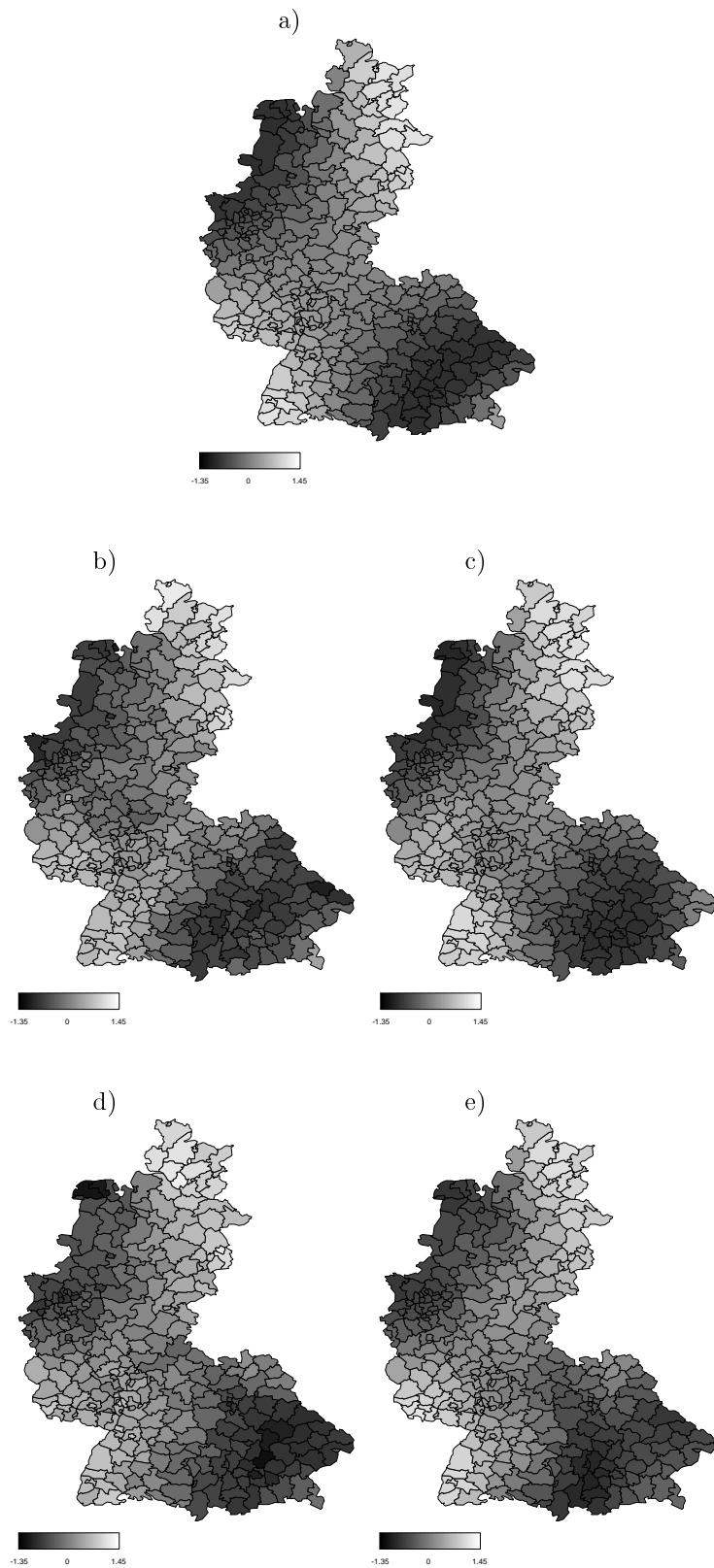


Figure 9: Spatial effects for the various model specifications; displayed are posterior mean estimates of run r , with r chosen such that MSE_r is the median of MSE_1, \dots, MSE_{100}

a) true function b) p.e.m., MRF, $r=65$, $MSE=0.041$ c) p.e.m., geospline, $r=83$, $MSE=0.021$
 d) P-spline model, MRF, $r=67$, $MSE=0.042$ e) P-spline model, geospline, $r=22$, $MSE=0.021$