Locally Adaptive Function Estimation for Binary Regression Models

Alexander Jerak and Stefan Lang

Department of Statistics, University of Munich.

April 14, 2003

Abstract

In this paper we present a nonparametric Bayesian approach for fitting unsmooth or highly oscillating functions in regression models with binary responses. The approach extends previous work by Lang et al. (2002) for Gaussian responses. Nonlinear functions are modelled by first or second order random walk priors with locally varying variances or smoothing parameters. Estimation is fully Bayesian and uses latent utility representations of binary regression models for efficient block sampling from the full conditionals of nonlinear functions.

Keywords: adaptive smoothing, forest health data, highly oscillating functions, MCMC, random walk priors, unsmooth functions, variable smoothing parameter.

1 Introduction

Nonparametric methods for fitting smooth curves, such as kernel, local or spline regression, are now widely available and accepted. However, these methods can have bad performance when estimating unsmooth functions which have jumps, edges, or which are highly oscillating. Two prominent approaches in nonparametric regression with Gaussian responses that adapt to such spatial heterogeneity are local regression with variable bandwidth (Fan and Gijbels, 1995) or wavelet shrinkage regression (Donoho and Johnstone, 1994). Currently, these methods are restricted to metrical responses and there is a clear lack of methodology and experience for non-Gaussian responses.

In this paper we present a nonparametric fully Bayesian method for fitting unsmooth and highly oscillating functions in regression models with *binary responses*. The approach extends recent work by Lang et al. (2002) for Gaussian responses. Our approach uses a two-stage prior for the unknown regression function. The first stage are first or second order random walk models as proposed in Fahrmeir and Lang (2001a) and Fahrmeir and Lang (2001b). The second stage consists of analogous smoothness priors for varying variances of the random walk model errors used in the first stage leading to locally adaptive dependent variances. The varying variances in our method correspond to variable smoothness parameters and make the prior more flexible for modelling functions with differing curvature. We compare our approach with random walk priors with a global variance as well as locally adaptive independent variances. The latter has been already used e.g. by Knorr-Held (1999) in the context of dynamic models.

Bayesian inference is based on latent *utility representations of binary regression models*, see Albert and Chib (1993) for probit models and Holmes and Knorr-Held (2003) for logit models. The advantage of augmenting the data by latent utilities is that the full conditionals of unknown parameters are Gaussian and efficient MCMC sampling schemes developed for Gaussian responses can be exploited.

The rest of this paper is organized as follows: Section 2 describes our Bayesian model for locally adaptive function estimation and gives details about Bayesian inference. Section 3 illustrates the performance of our approach by selected results from an extensive simulation study. In Section 4 the practicability is demonstrated by a complex application on forest health data. The final section 5 summarizes the paper and highlights directions for future research.

2 Model specification and Bayesian inference

2.1 Binary response models

Consider regression situations, where observations (y_t, z_t), $t = 1, \ldots, T$, on a binary response y and covariates z are given. The most widely used models for binary data are logit or probit models. Given the covariates the responses y_t are binomially distributed, i.e. $y_t|z_t \sim B(1, \pi_i)$ with the probability of success $\pi_t = P(y_t = 1|z_t) = E(y_t|z_t)$ being modeled as

$$\pi_t = \frac{\exp(\eta_t)}{1 + \exp(\eta_t)}$$

for logit models or

$$\pi_t = \Phi(\eta_t)$$

for probit models. Here, η_t is the predictor that models the influence of the covariates. With a linear predictor

$$\eta_t = z_t'\beta \tag{1}$$

one gets parametric models. In many practical situations, as in our application on forest health data, the assumption of linear effects of the covariates on the predictor is too restrictive. Suppose the covariates z_t are divided into a vector of metrical covariates $x_t = (x_{t1}, \ldots, x_{tp})'$ whose influence is assumed to be possibly nonlinear, and a vector of categorical covariates $w_t = (w_{t1}, \ldots, w_{tq})'$. Then, we replace the simple linear predictor (1) by the semiparametric additive predictor

$$\eta_t = f_1(x_{t1}) + \cdots + f_p(x_{tp}) + w_t'\gamma,$$

were we assume possibly nonlinear effects $f_1, \ldots, f_p$ for the metrical covariates. In this paper, the primary focus is on modelling functions with discontinuities and/or differing curvature. We will discuss appropriate prior specifications for functions of this kind in the next section.

For Bayesian inference, it is quite useful to express binary regression models in terms of latent utilities (e.g. Fahrmeir and Tutz (2001)). Introducing the latent utilities

$$U_t = \eta_t + \epsilon_t \tag{2}$$

with i.i.d. errors ϵ_t, we define $y_t = 1$ if $U_t > 0$ and $y_t = 0$ if $U_t < 0$. The assumption $\epsilon_t \sim N(0,1)$ yields a probit model. A logit model is obtained by assuming $\epsilon_t \sim N(0, \lambda_t)$ with $\lambda_t = 4\psi_i^2$, where ψ_t follows a Kolmogorov-Smirnov distribution (Devroye (1986)). Hence, ϵ_t is a scale mixture of normal form with a marginal logistic distribution (Andrews and Mallows (1974)). Note, that a logit model could be (well) approximated by assuming a t-distribution for the ϵ_t's with a certain degree ν of freedom. A t-distribution may again be expressed as a scale mixture of normals with $\lambda_t \sim IG(\nu/2, \nu/2)$. An approximative logit model is then obtained with $\nu = 8$ (Albert and Chib (1993)).

2.2 Prior models

For Bayesian inference, the unknown functions f_j, $j = 1, \ldots, p$, or more exactly the corresponding vectors of function evaluations, are considered as random and must be supplemented by appropriate prior distributions. In the following, let f denote one of the unknown functions f_j and x the corresponding covariate. Let $x_{(1)} < x_{(2)} < \cdots < x_{(S)}$ denote the ordered sequence of observed covariate values. Define $f_s := f(x_{(s)})$, $s = 1, \ldots, S$, and let $f = (f_1, \ldots, f_S)'$ be the vector of function evaluations.

Assuming equidistant covariate values, a common prior for a smooth function f is a first or second order random walk model

$$f_s = f_{s-1} + u_s \quad (RW1) \quad \text{or} \quad f_s = 2f_{s-1} - f_{s-2} + u_s \quad (RW2)$$

with Gaussian errors $u_s \sim N(0, \tau^2)$ and diffuse priors $f_1 \propto constant$, or f_1 and $f_2 \propto constant$, for initial values, respectively. Generalizations to situations with non-equally spaced observations are given in Fahrmeir and Lang

(2001a). For instance, a second order random walk could be generalized to

$$f_s = \left(1 + \frac{\delta_s}{\delta_{s-1}}\right) f_{s-1} - \frac{\delta_s}{\delta_{s-1}} f_{s-2} + u_s,$$

where $u(t) \sim N(0; w_s\tau^2)$, $\delta_s = x_{(s)} - x_{(s-1)}$ and w_s is an appropriate weight, e.g. $w_s = \delta_s$. To avoid notational confusions, in the following we restrict the presentation to the case of equally spaced observations. A generalization to the more general case is straightforward and supported by our software.

The amount of smoothness is controlled by the variance parameter τ^2 which corresponds to the smoothing parameter in a classical approach. To estimate the amount of smoothness simultaneously with the unknown function a highly dispersed inverse gamma prior $\tau^2 \sim IG(a, b)$ is assigned in a further stage of the hierarchy.

For unsmooth or highly oscillating functions, as primarily considered in this paper, the assumption of a global variance or smoothing parameter is however not appropriate. We therefore replace the constant variance τ^2 by *locally adaptive* variances τ_s^2. In the following we will discuss approaches with both *stochastic dependent* and *independent* τ_s^2.

We start with the stochastic *dependent* variant. Following Lang et al. (2002) we set $\tau_s^2 = \exp(h_s)$. For the parameters h_s we add a second smoothness prior in the form of first or second order random walks for $h = (h_d, \dots, h_S)$, i.e.

$$h_s = h_{s-1} + v_s \quad \text{or} \quad h_s = 2h_{s-1} - h_{s-2} + v_s, \quad v_s \sim N(0, \sigma^2). \tag{3}$$

The index d depends on the choice of the prior for f, for a RW(1) $d = 2$ and for a RW(2) $d = 3$. Once again, a highly dispersed inverse gamma prior is assigned to the variance parameter σ^2.

Alternatively we may assume *independent* local variances τ_s^2 instead of dependent variances, which may be particularly useful for functions with discontinuities. Assuming i.i.d. inverse gamma priors

$$\tau_s^2 \sim IG(v/2, v/2) \tag{4}$$

the marginal distribution of the errors is a Student distribution with v degrees of freedom. The case $v = 1$ of a Cauchy distribution is of special interest as a robust prior and is used for the rest of this paper.

We conclude this section with a few additional prior assumptions:

1. For the fixed effects parameters β we assume independent diffuse priors, i.e. $\beta_j \propto const$, $j = 1, \dots, q$.

2. For given covariates and parameters observations y_t are conditionally independent.

3. Priors for the functions f_j, $j = 1, \dots, p$, and fixed effects are mutually independent.

2.3 Bayesian inference via MCMC

For binary regression models a useful and very efficient sampling scheme can be developed on the basis of the latent variables representation with Gaussian errors for probit models and a scale mixture of normal form for logit models defined in (2). We start by describing inference for models with locally dependent variances $\tau_s^2 = \exp(h_s)$ and random walk priors (3) for h_s. Let $f_j := (f_{j1}, \dots, f_{jS_j})$ be the vector of function evaluations of the j-th covariate x_j. Define h_j as the corresponding vector of variance parameters and let σ_j^2 be the j-th variance parameters. Bayesian inference is based on the posterior augmented by the latent variables $U_t = (U_1, \dots, U_T)$

$$p(\dots, f_j, h_j, \tau_j^2, \sigma_j^2, \dots, \beta, U|y) \propto p(y|U)p(U|f_1, \dots, f_p, \beta) \prod_{j=1}^{p} \{p(f_j|h_j)p(h_j|\sigma_j^2)p(\sigma_j^2)\},$$

with $p(y|U) = \prod_t p(y_t|U_t)$. The conditional likelihood $p(y_t|U_t)$ is given by

$$p(y_t|U_t) = I(U_t > 0)I(y_t = 1) + I(U_t < 0)I(y_t = 0), \quad (5)$$

due to the fact that $p(Y_t|U_t)$ is one if U_t obeys the constraint imposed by the observed value of Y_t. MCMC sampling is based on successive drawings from the full conditionals of U_t, $t = 1, \dots, T$, f_j, $j = 1, \dots, p$, h_j and σ_j^2. It turns out, that updating of the variance parameters vector h_j in one step is not feasible because of too small acceptance rates. As a remedy the parameter vector h_j must be further divided into smaller blocks (usually of size 10-20), see Lang et al. (2002) for details on updating the variance parameters.

The full conditionals for the U_t's are truncated normals. For $y_t = 1$ we obtain

$$U_t|\cdot \sim N(\eta_t, \lambda_t)I(U_t > 0)$$

and for $y_t = 0$ we get

$$U_t|\cdot \sim N(\eta_t, \lambda_t)I(U_t < 0)$$

with $\lambda_t = 1$ for probit models. Drawing random numbers form a truncated normal distribution poses no further problems, see e.g. Robert (1995). For logit models additional drawings from the conditional distributions of λ_t are necessary. Although the distribution has no standard form, sampling may be obtained by Metropolis-Hastings steps with the prior distribution for λ_t as a proposal (Holmes and Knorr-Held (2003)), i.e. draw a random number ψ_t from a Kolmogorov Smirnov distribution and propose $\lambda_t^{prop} = 4\psi^2$ as the new state of the Markov chain. The proposed new value is then accepted with probability

$$\alpha = \left(\frac{\lambda_t}{\lambda_t^{prop}}\right)^{0.5} \exp\left(\frac{1}{2}(U_t - \eta_t)^2 \left(\frac{1}{\lambda_t} - \frac{1}{\lambda_t^{prop}}\right)\right),$$

where λ_t is the current state of the chain. If a t-distribution is assumed for the errors, the full conditionals for λ_t are inverse Gamma distributions with parameters $\nu/2+1/2$ and $\nu/2+(U_t-\eta_t)^2/2$. From a computational point of view we prefer probit models because updates of the full conditionals of the latent utilities are (slightly) faster compared to logit models or the t-link.

The advantage of augmenting the posterior by the latent variables is that the full conditionals for functions f_j become Gaussian, allowing the usage of the sampling schemes developed for Gaussian responses in Lang et al. (2002) with only minor changes. Updating of f_j, h_j and σ_j^2, $j=1,\dots,p$, can be done exactly as described in Lang et al. (2002) using the current values of the latent utilities as (pseudo) responses.

If we assume locally independent variances $\tau_j^2 = (\tau_{j1}^2,\dots,\tau_{jS_j}^2)$ with priors (4), the sampling scheme facilitates considerably. Updating of the variance parameters is straightforward because the full conditionals for τ_{js}^2, $s=1,\dots,S_j$ are inverse Gamma distributions with parameters $\nu/2+1/2$ and $\nu/2+(f_s-f_{s-1})^2/2$ if a first order random walk is assumed for f_j. In the case of a second order random walk for f_j, the parameters are $\nu/2+1/2$ and $\nu/2+(f_s-2f_{s-1}+f_{s-2})^2/2$. Updating of the remaining parameters is exactly the same as described above for locally dependent variances.

3 Simulation studies

To illustrate the performance of our locally adaptive approaches we carried out two simulation studies for binomial probit models with different settings for the true regression function. Binomial response vectors $y=(y_1,\dots,y_T)$ were generated by setting $\eta_t = f(x_t)$ and drawing $B(n,\Phi(\eta_t))$ distributed random variables y_t, $t=1,\dots,T$. To assess the dependence of results on the number of observations we used both $n=1$ and $n=3$. For the true regression function f we considered the two cases depicted in Figure 1. In the first case ($T=250$) a discontinuous step function for f and in the second case ($T=400$) a regression function characterized by differing curvature and medium spatial variability taken from Ruppert and Carroll (2000) was used. For each of the two cases we generated 250 replications and applied the three approaches with global variances, locally dependent and independent variances decsribed in Section 2 to each replication.

A similar simulation study based on logit models rather than probit models shows virtually identical results. Therefore, and to keep the paper in reasonable length, results for logit models are not presented.

3.1 Regression function with discontinuities

Facing a regression function with discontinuities, the best results for all approaches were usually obtained by using a RW1 prior for the regression function f and, in case of the approach with locally dependent variances, a

RW1 prior for the variance function h. We therefore restrict the presentation to these cases and denote them in the following with RW1 (global approach), TRW1 (locally independent variances) and RW1VRW1 (locally dependent variances).

Figures 2 and 3 display for $n = 1$ and $n = 3$ boxplots of $\ln(MSE) = \ln(1/T \sum_{t=1}^{T}(f(x_t) - \hat{f}_t(x_t))^2)$ and regression function estimates averaged over the 250 replications for the various estimators. The respective panels (a) in Figures 4 and 5 show variance function estimates, again averaged over the 250 replications. Panels (b) show the coverage probabilities of pointwise credible intervals for a nominal level of 95%. In a Bayesian approach based on MCMC simulations, pointwise credible intervals are simply obtained by computing the respective quantiles of the sampled function evaluations. From Figures 2 - 5 we can draw the following conclusions:

- For $n = 1$ the three approaches show similar performance in terms of MSE and bias. Inspection of individual estimates reveals that only one observation per covariate value is not enough to recover the underlying step function satisfactorily.
- With $n = 3$ observations, the true curve can be recovered satisfactorily. As expected, the best results in terms of MSE and bias are obtained with TRW1, i.e. locally adaptive independent variances.
- The jumps in the regression function are best reflected in the variance functions for TRW1. For the approach with locally dependent variances RW1VRW1 the variance functions show only for $n = 3$ observations a significant increase of variances at all jumps. The increase of variances is, of course, less pronounced for RW1VRW1 as for TRW1.
- Even for $n = 1$, for both approaches with adaptive variances the coverage rates are closer to the nominal level than for the approach with a global variance. The approach TRW1 reveals a dramatic improvement of coverage rates at the jumps compared to RW1 and RW1VRW1.

3.2 Regression function with differing curvature

In contrast to a regression function with discontinuities, the best results for all approaches were usually obtained by using a RW(2) prior for the regression function f and, in case of the approach with locally dependent variances, a RW(1) prior for the variance function h. We therefore restrict the presentation to these cases and denote them in the following with RW2 (global approach), TRW2 (locally independent variances) and RW2VRW1 (locally dependent variances).

Figures 6 and 7 display boxplots of $\ln(MSE)$ and regression function estimates averaged over the 250 replications. The average variance function estimates are depicted in panel (a) of Figures 8 and 9. The coverage rates of pointwise credible intervals are shown in panels (b). The results shown in Figures 6 - 9 can be interpreted as follows:

- In contrast to functions with discontinuities, satisfactory results are now obtained even for $n = 1$.
- Not surprisingly, the approach RW2VRW1 with locally dependent variances clearly outperforms the global approach RW2 and the locally independent approach TRW2. Most striking is the severe bias for $n = 1$ obtained with TRW2 at the local minima and maxima of the curve. A possible explanation might be that the approach with local independent variances is too flexible in situations where the true probabilities of success are close to one or zero (as is the case at the minima and maxima of the curve).
- The decrease of spatial variability in the regression function is accurately reflected by the variance functions obtained from RW2VRW1 while for TRW2 the variances stay more or less constant at the level of the approach with global variance RW2.
- The coverage rates for all three estimators are generally close to the nominal level. For TRW2 and $n = 1$, however, the coverage is clearly below the nominal level near the minima and maxima of the true curve. The main reason is the strong bias in this area.

4 Application to forest health data

In this section we demonstrate the practicability of our methods by an application to forest health data. We analyse the influence of calendar time, age, canopy density CP and location L on the health state of trees (y=1 for a damaged tree and y=0 otherwise). Data have been collected in yearly forest damage inventories carried out in the forest district of Rothenbuch in northern Bavaria from 1983 to 2001. 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 12. A detailed data description can be found in Göttlein and Pruscha (1996).

We used a binary probit model with predictor

$$y_{it} = f_1(t) + f_2(age_i) + f_{spat}(L_i) + \beta \cdot CP_i$$

for tree i, $i = 1, \ldots, 80$ and year t, $t = 1983, \ldots, 2001$. Here, age_i is the age of the tree in years at the beginning of the observation period, L_i is the location of tree i, and CP_i is the canopy density at the stand in percent (0%,10%,20%,...,100%). Preliminary examination of the data reveal that the effect of canopy density is linear. Therefore CP is included as a usual fixed effect with a diffuse prior for the regression coefficient.

Although this is only a demonstrating example, it is important to consider possible spatial heterogeneity of the data for a realistic modelling approach which captures the most important features of the data. For that reason we included a spatial effect f_{spat}. We assigned a Markov random field prior (Besag et al. (1991)), with the neighbourhood ∂_L of trees including all trees L' with euclidian distance $d(L, L') \leq 1.2$ km, see also Fahrmeir and Lang (2001b). Our model is an example for a regression model with *geoadditive predictor* (Kammann and Wand (2003)) and demonstrates one of the main advantages of Bayesian inference for semiparametric regression based on MCMC simulation: models can be easily extended to more complex formulations.

Figure 10 (a) and (b) shows estimated functions f_1 and f_2 based on a first order random walk with a global variance. The respective results based on a second order random walk are depicted in Figure 11. The estimated spatial effects can be found in Figure 12. We see that trees recover after the bad years around 1986, but after 1992 health status declines to a lower level again. As we might have expected, younger trees are in healthier state than the older ones. Note also, that the incorporation of the spatial effect into the model is quite important since the estimated effect suggests considerable spatial heterogeneity.

Starting from these two models, experiments with our spatially adaptive random walk priors gave evidence for a jump of the age effect around age 20 and hints for a smoothly varying variance of the time trend. We therefore replaced the global variance of the age effect by locally independent variances (4) and the global variance of the time trend by locally dependent variances (3) with a first order random walk for h_s. Figure 10 (c) and (d) show results based on first order random walks for f_1 and f_2 and Figure 11 (c) and (d) results based on second order random walks for f_1 and f_2. The respective figures (e) and (f) display the estimated locally varying variance functions. Results for the spatial effect remain almost unchanged compared to our basis models and are therefore not replicated.

In terms of the DIC (Spiegelhalter et al. (2002)) both estimates with varying variances are clearly superior to the models with global variances. As could have been expected, the estimated jump for the age effect is more pronounced with a first order random walk rather than a second order random walk for f_2. The smallest DIC is however obtained by the model based on second order random walks for f_1 and f_2 although the differences are rather small.

5 Conclusions

This paper presents a practical approach for fitting highly oscillating or unsmooth functions in binary regression models. The simulation study in

Section 3 suggests that for highly oscillating functions the approach with locally dependent variances performs superior to locally independent variances and simple random walk models with a global variance. For jump functions results are superior with locally independent variances.

For probit models, our approach can be extended to models mit multicategorical responses by using similar latent utility respresentations as for binary responses, see e.g. Fahrmeir and Lang (2001b). For multinomial logit models, such representations can not be used for facilitating MCMC inference. In this case, the development of a direct sampling scheme which is not based on latent utilities is necessary. A possible approach could be based on iteratively weighted least squares proposals for the nonlinear functions f_j as proposed for generalized linear models by Gamerman (1997). We consider to extend our models to the multicategorical case in future research.

Another aspect for future research concerns model choice. The introduction of locally adaptive function estimates complicates model choice considerably, because one has to decide not only whether a covariate should be included into the model or not, but also how the covariate effect should be modeled. In our application we used the DIC as a goodness of fit measure. The drawback of model choice via the DIC is that only a limited number of models can be tested. For the future, we plan to develop Bayesian inference techniques that allow estimation and model choice (to some extent) simultaneously.

References

Albert, J. and Chib, S., 1993: Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88, 669-679.

Andrews, D.F. and Mallows, C.L., 1974: Scale mixtures of normal distributions. *Journal of the Royal Statistical Society B*, 36, 99-102.

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

Devroye, L., 1986: *Non-Uniform Random Variate Generation.* Springer-Verlag, New York.

Donoho, D. L. and Johnstone, I. M., 1993: Ideal spatial adaption by wavelet shrinkage. *Biometrika*, 81, 425-455.

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 (Applied Statistics)*, 50, 201-220.

Fahrmeir, L. and Lang, S., 2001: Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data. *Annals of the Institute of Statistical Mathematics*, 53, 10-30

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

Fan, J. and Gijbels, I., 1995: Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption. *Journal of the Royal Statistical Society B*, 57, 371-394.

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

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

Holmes, C.C., and Knorr-Held, L., 2003: Efficient Simulation of Bayesian Logistic Regression Models. Discussion paper 306, SFB 386, Department of Statistics, University of Munich.

Kamman, E. E. and Wand, M. P., 2003: Geoadditive Models. *Journal of the Royal Statistical Society C (Appl. Stat.)*, to appear.

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

Lang, S., Fronk, E.-M., Fahrmeir, L., 2002: Function estimation with locally adaptive dynamic models. *Computational Statistics*, 17, 479-500.

Ruppert, D. and Carroll, R. J., 2000: Spatially adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics*, 42, 205-223.

Robert, C.P., 1995: Simulation of truncated normal variables. *Statistics and Computing*, 5, 121-125.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A., 2002: Bayesian measures of model complexity and fit., *Journal of the Royal Statistical Society B*, 65, 583 - 639.

Address:
Alexander Jerak, Department of Statistics, University of Munich, Ludwigstr. 33, 80539 Munich, Germany. email: jerak@stat.uni-muenchen.de

Stefan Lang*, Department of Statistics, University of Munich, Ludwigstr. 33, 80539 Munich, Germany. email: lang@stat.uni-muenchen.de
*corresponding author

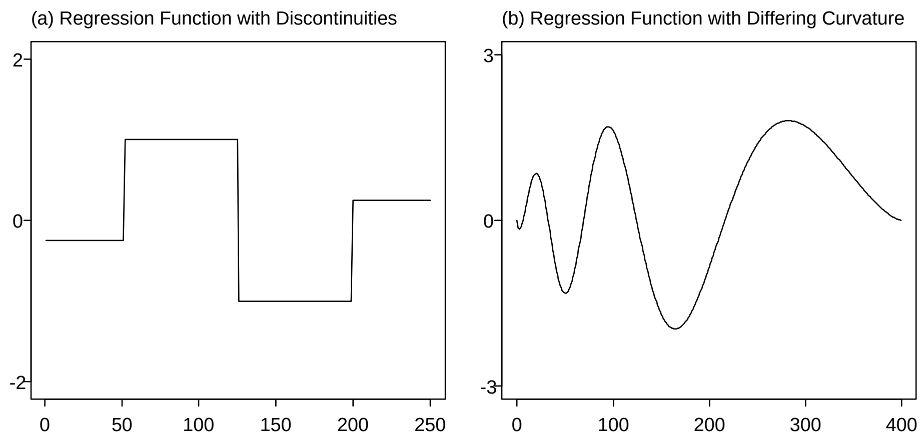

Figure 1: True regression functions used in simulation studies.

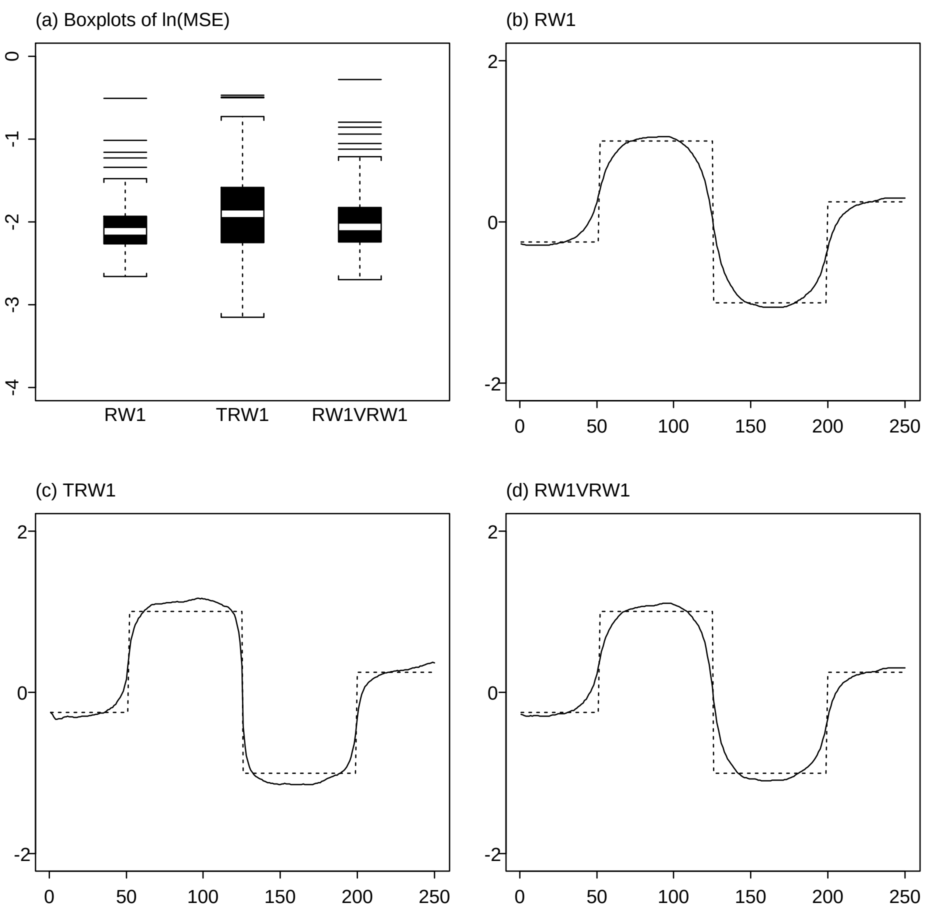

Figure 2: Simulation results for regression function with discontinuities and one observation on each design point. (a) Boxplots of ln(MSE). (b)-(d) Averaged posterior mean estimates (—) together with true function $(\cdots)$.

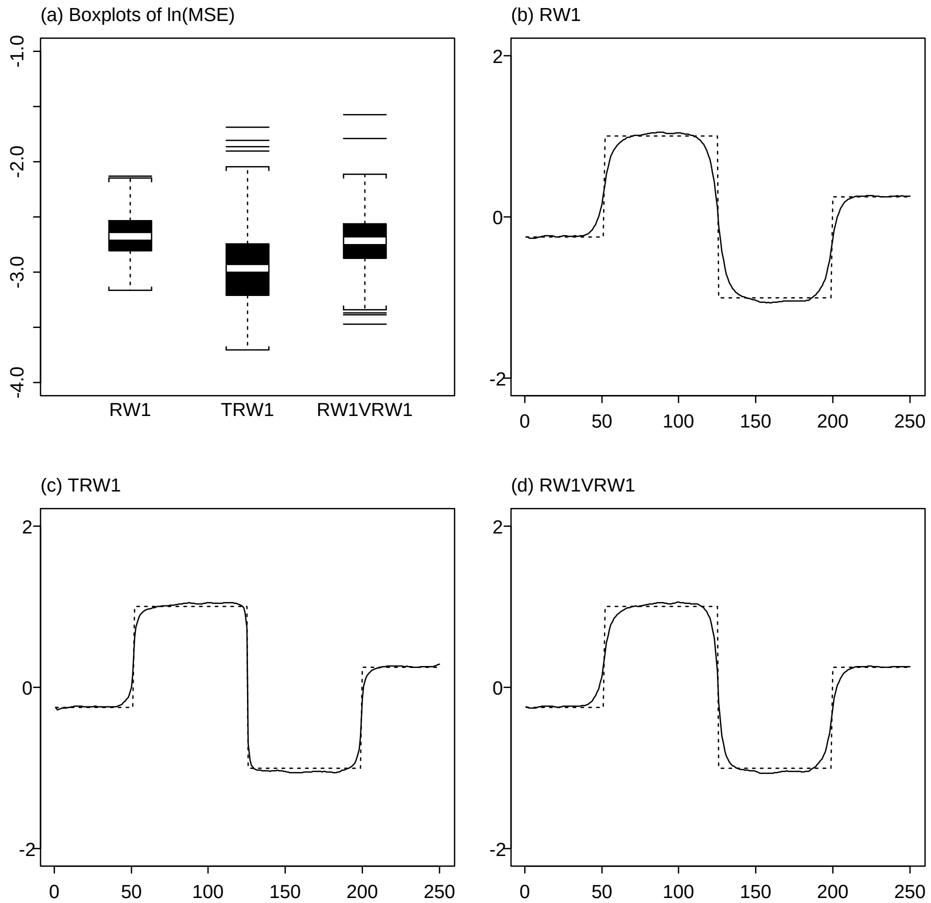

Figure 3: Simulation results for regression function with discontinuities and three observations on each design point. (a) Boxplots of ln(MSE). (b)-(d) Averaged posterior mean estimates (—) together with true function ($\cdots$).

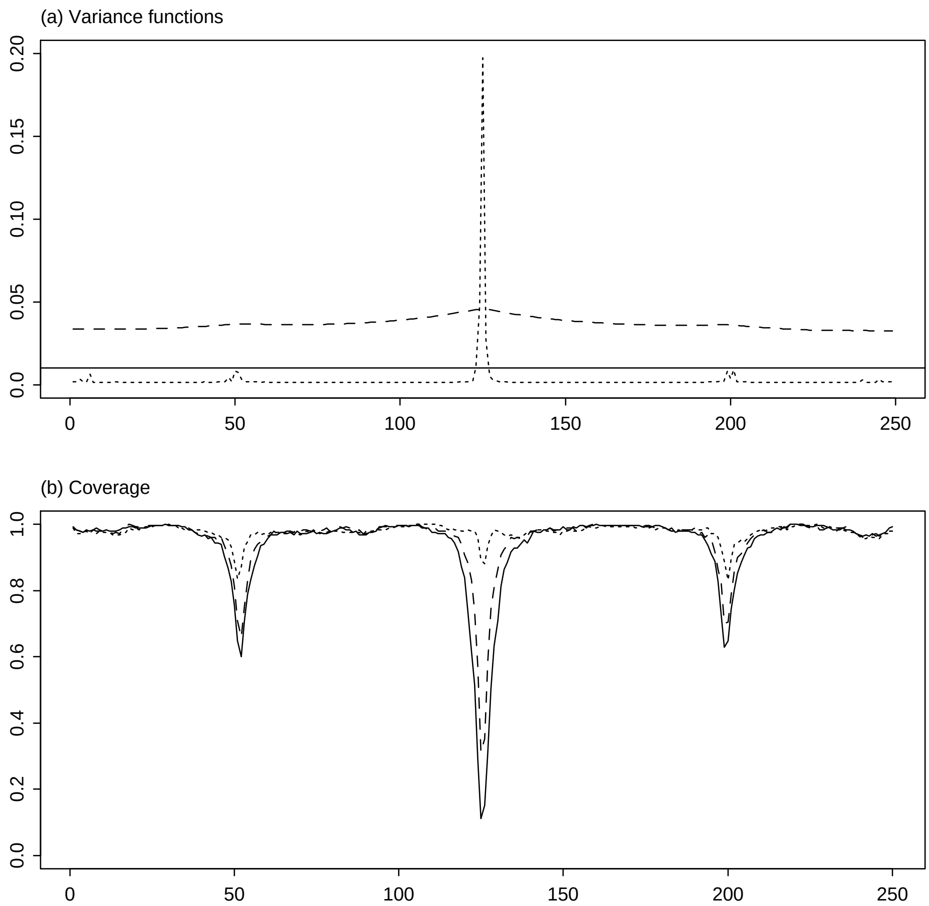

Figure 4: Simulation results for regression function with discontinuities and one observation on each design point in the approach with global variance (—), locally dependent variances (- -) and locally independent variances ($\cdots$). (a) Averaged posterior median estimates for variance functions. (b) Coverage of pointwise 95 % credible intervals.

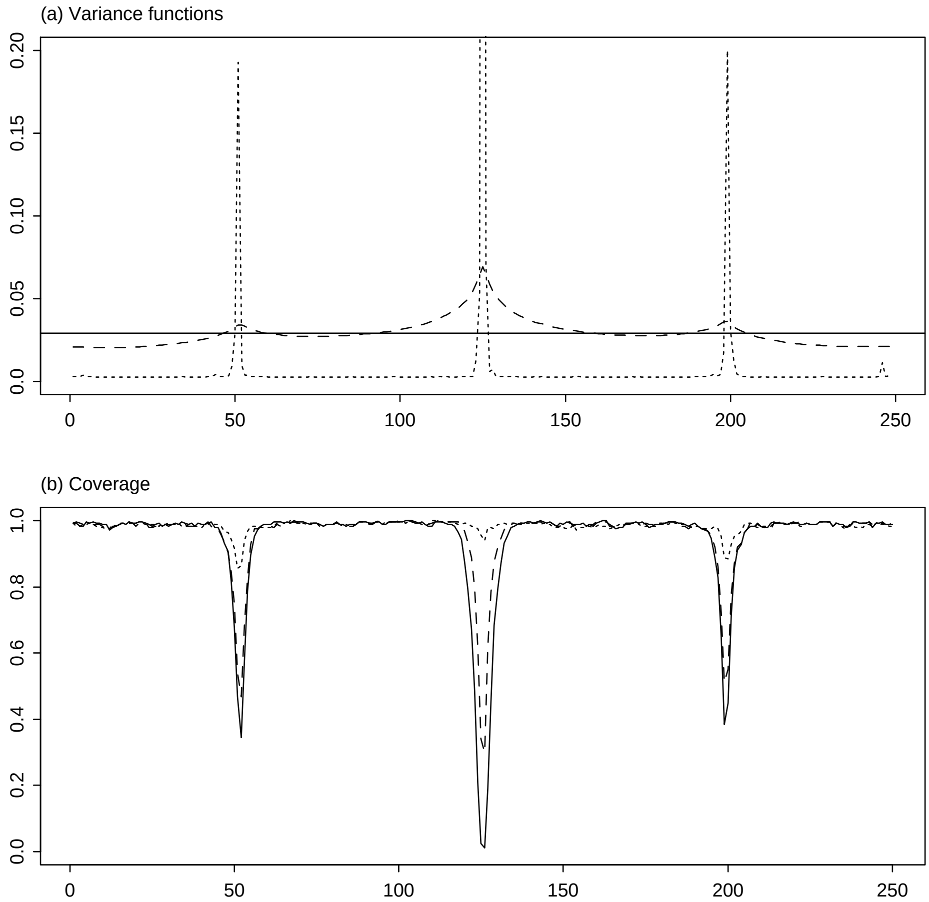

Figure 5: Simulation results for regression function with discontinuities and three observations on each design point in approach with global variance (—), locally dependent variances (- -) and locally independent variances ($\cdots$). (a) Averaged posterior median estimates for variance functions. (b) Coverage of pointwise 95 % credible intervals.

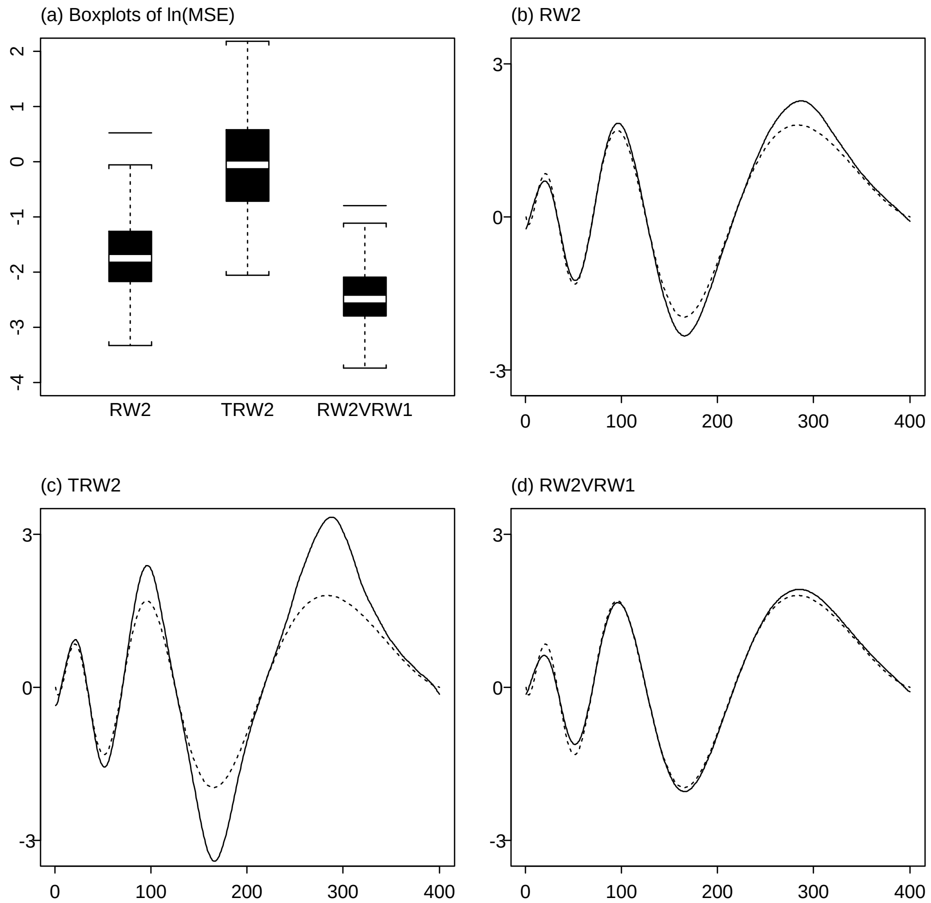

Figure 6: Simulation results for regression function with differing curvature and one observation on each design point. (a) Boxplots of ln(MSE). (b)-(d) Averaged posterior mean estimates (—) together with true function $(\cdots)$.

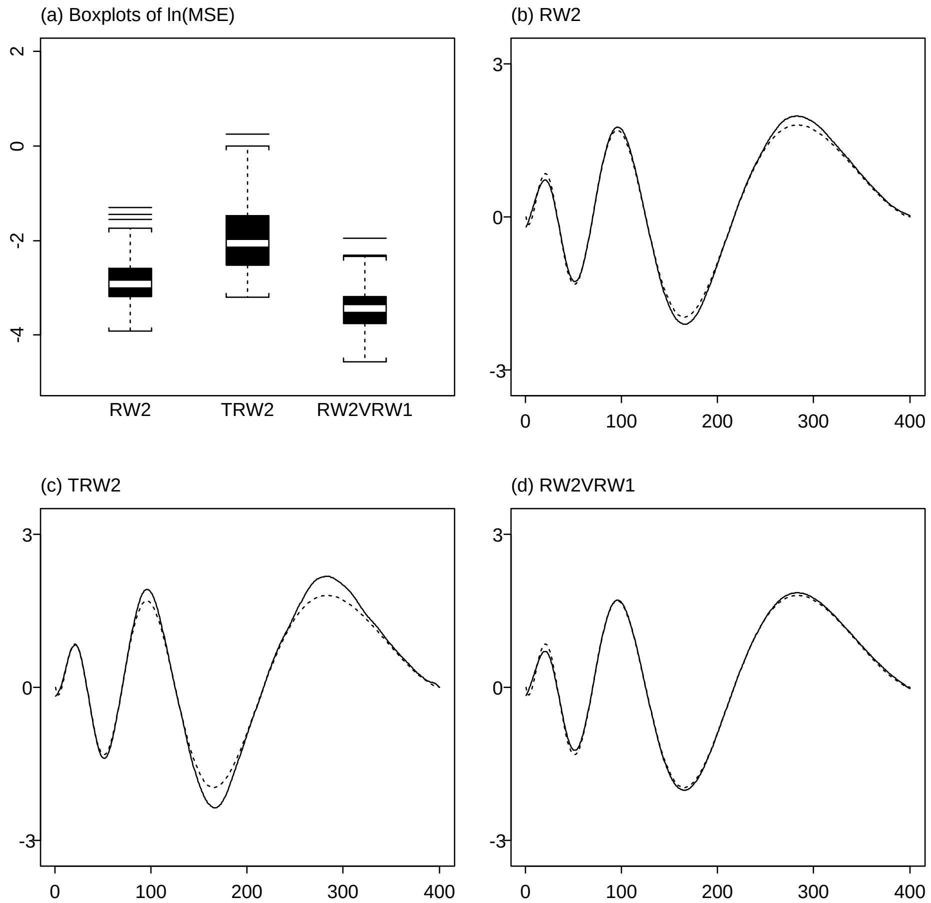

Figure 7: Simulation results for regression function with differing curvature and three observations on each design point. (a) Boxplots of ln(MSE). (b)-(d) Averaged posterior mean estimates (—) together with true function ($\cdots$).

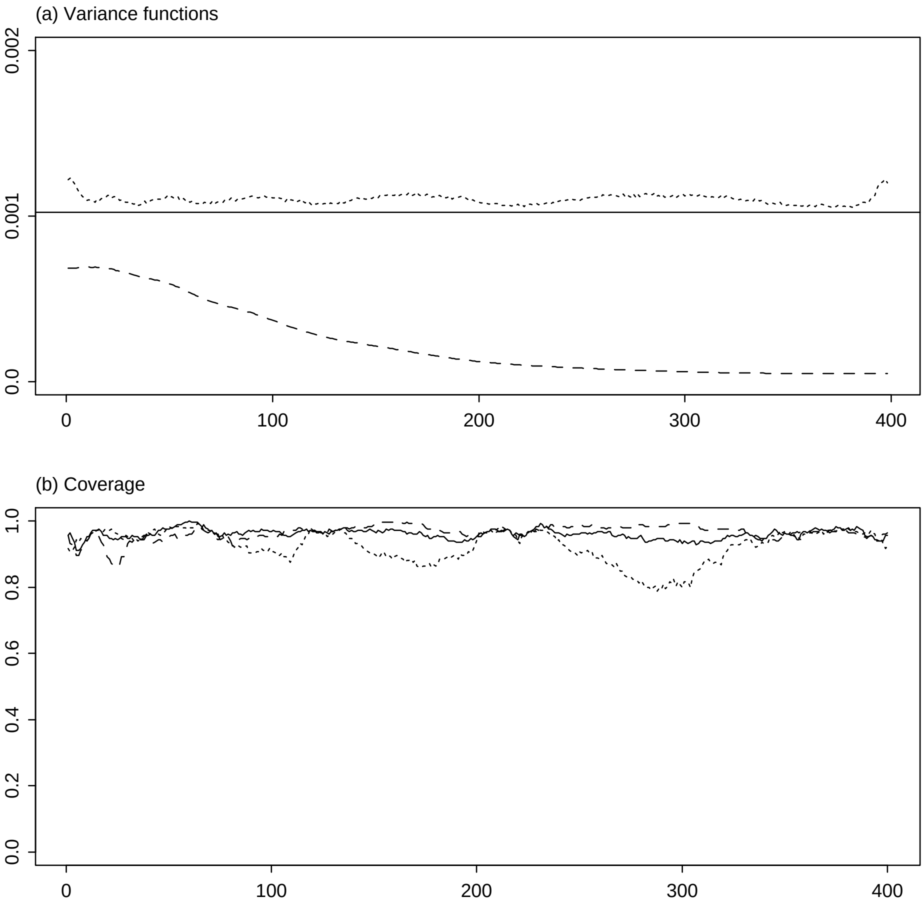

Figure 8: Simulation results for regression function with differing curvature and one observation on each design point in approach with global variance (—), locally dependent variances (- -) and locally independent variances ($\cdots$). (a) Averaged posterior median estimates for variance functions. (b) Coverage of pointwise 95 % credible intervals.

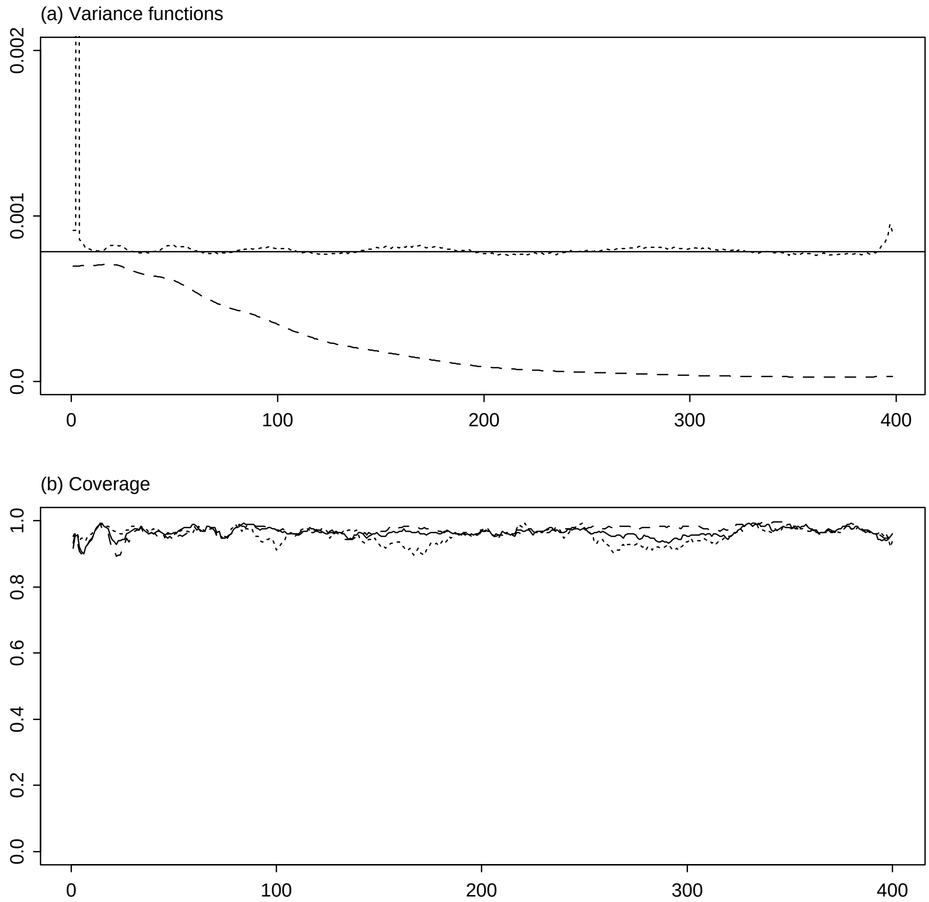

Figure 9: Simulation results for regression function with differing curvature and three observations on each design point in approach with global variance (—), locally dependent variances (- -) and locally independent variances ($\cdots$). (a) Averaged posterior median estimates for variance functions. (b) Coverage of pointwise 95 % credible intervals.

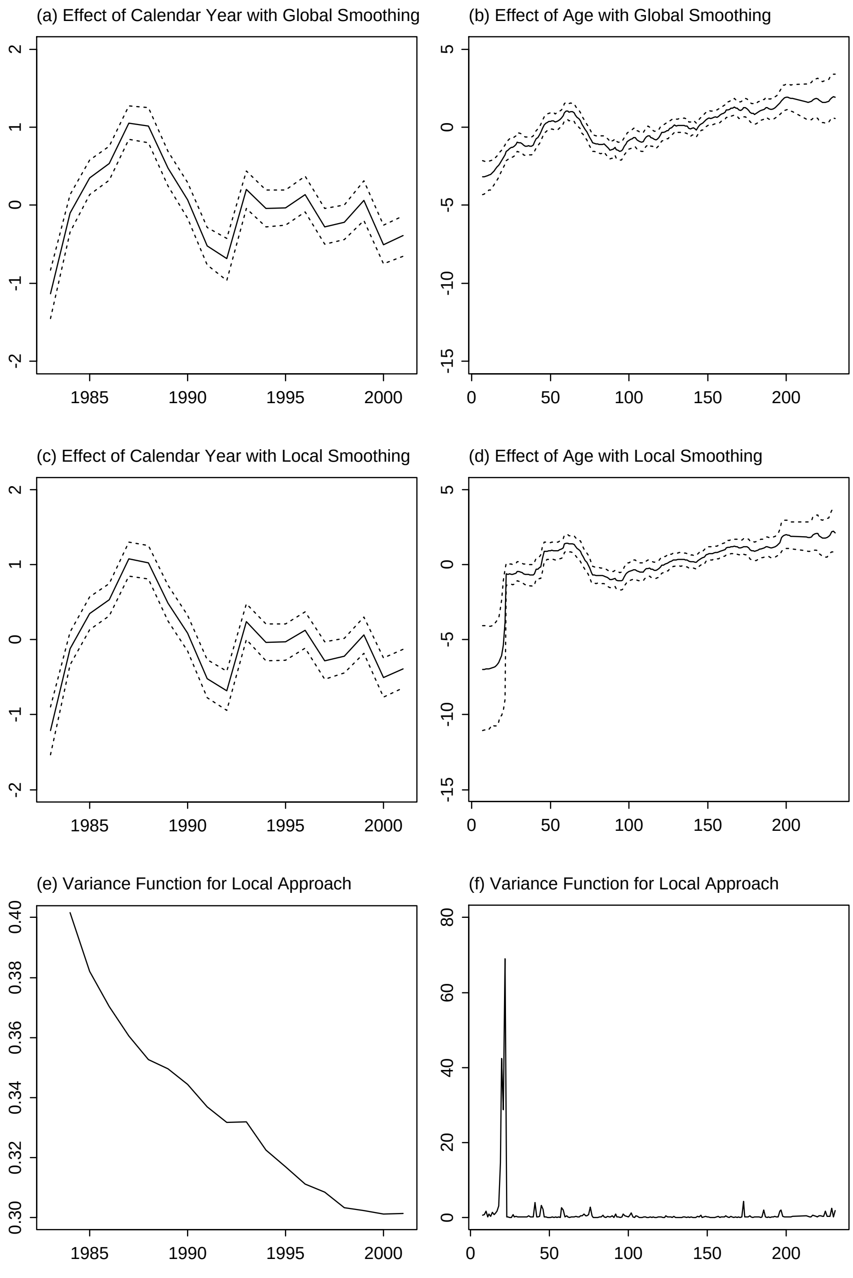

Figure 10: Results for forest health data. (a)-(b) Posterior mean estimates (—) and 80 % credible intervals ($\cdots$) in global approach with RW1 for both calendar year and age effect. (c)-(d) Posterior mean estimates (—) and 80 % credible intervals ($\cdots$) in local approach with RW1VRW1 for calendar year and TRW1 for age effect. (e)-(f) Posterior median estimates for variance functions in local approach.

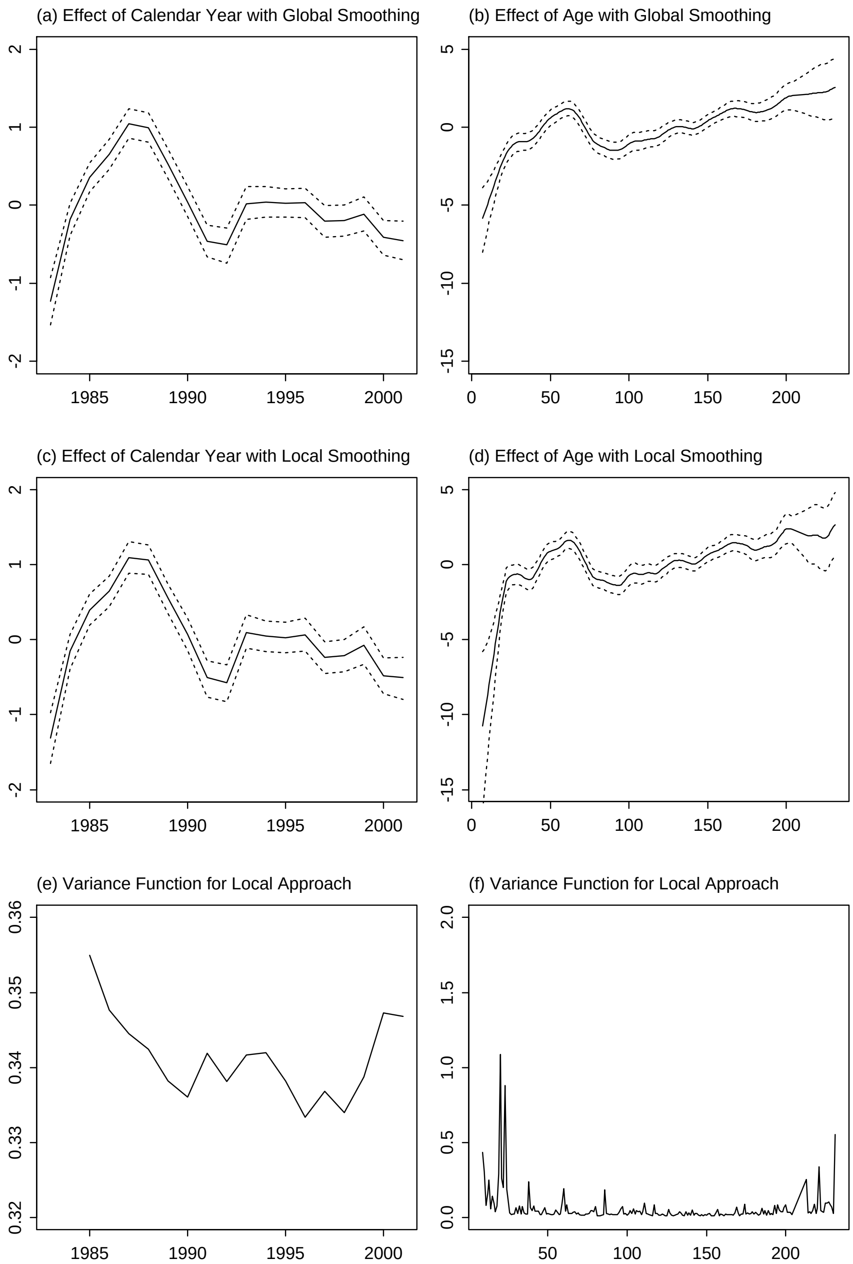

Figure 11: Results for forest health data. (a)-(b) Posterior mean estimates (—) and 80 % credible intervals ($\cdots$) in global approach with RW2 for both calendar year and age effect. (c)-(d) Posterior mean estimates (—) and 80 % credible intervals ($\cdots$) in local approach with RW2VRW1 for calendar year and TRW2 for age effect. (e)-(f) Posterior median estimates for variance functions in local approach.

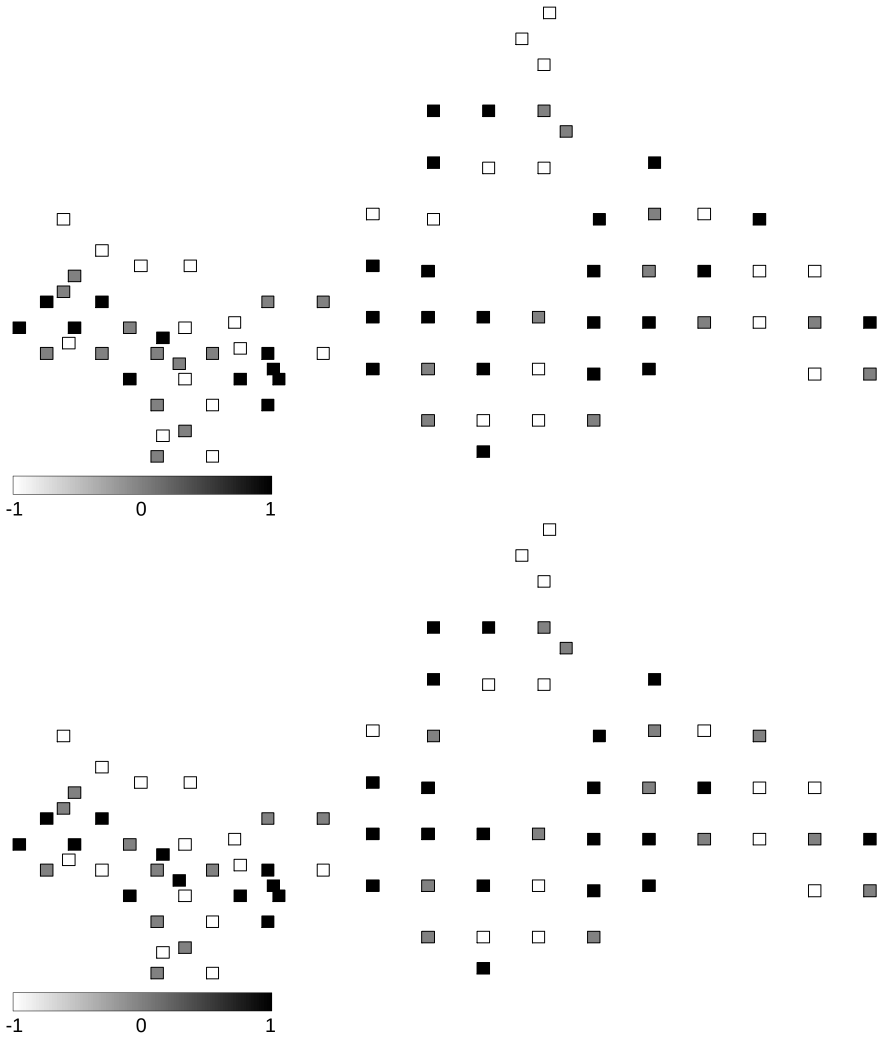

Figure 12: Results for forest health data. Posterior probabilities (based on a nominal level of 80%) for spatial effects of the two models with global variances. The top panel corresponds to the model with first order random walks and the bottom panel to the model with second order random walks for f_1 and f_2. Black spots indicate a positive, white spots a negative and grey spots a non-significant effect.