Open access • Journal Article • DOI:10.1111/J.1541-0420.2008.01112.X

Variable Selection and Model Choice in Geoadditive Regression Models

Thomas Kneib, Torsten Hothorn, Gerhard Tutz

Institutions: Ludwig Maximilian University of Munich

Published on: 01 Jun 2009 - Biometrics (Blackwell Publishing Inc)

Topics: Regression analysis, Random effects model, Covariate, Bivariate analysis and Logistic regression

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Thomas Kneib, Torsten Hothorn & Gerhard Tutz

Variable Selection and Model Choice in Geoadditive Regression Models

Technical Report Number 003, 2007
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Variable Selection and Model Choice in Geoadditive Regression Models

Thomas Kneib,[1,*] **Torsten Hothorn**[1] **and Gerhard Tutz**[1]

[1] Institut für Statistik
Ludwig-Maximilians-Universität München
Ludwigstraße 33, D-80539 München, Germany

Summary.

Model choice and variable selection are issues of major concern in practical regression analyses. We propose a boosting procedure that facilitates both tasks in a class of complex geoadditive regression models comprising spatial effects, nonparametric effects of continuous covariates, interaction surfaces, random effects, and varying coefficient terms. The major modelling component are penalized splines and their bivariate tensor product extensions. All smooth model terms are represented as the sum of a parametric component and a remaining smooth component with one degree of freedom to obtain a fair comparison between all model terms. A generic representation of the geoadditive model allows to devise a general boosting algorithm that implements automatic model choice and variable selection. We demonstrate the versatility of our approach with two examples: a geoadditive Poisson regression model for species counts in habitat suitability analyses and a geoadditive logit model for the analysis of forest health.

Key words: bivariate smoothing, boosting, functional gradient, penalised splines, random effects, space-varying effects

---

* *email:* thomas.kneib@stat.uni-muenchen.de

1. Introduction

Generalized linear models (GLM) have become one of the standard tools for analyzing the impact of covariates on possibly non-Gaussian response variables. A crucial question in setting up a GLM for a particular application is the choice of an appropriate subset of the set of available covariates, i.e., variable selection. In addition, one has to determine how to model the covariate effects, a task we will refer to as model choice in the following. While variable selection and model choice issues are already complicated in linear models and GLMs and still receive considerable attention in the statistical literature (see,e.g., George, 2000; Fan and Li, 2001; Zou and Hastie, 2005; Bühlmann, 2006, for recent approaches and discussion), they become even more challenging in geoadditive regression models including nonparametric effects of continuous covariates, spatial effects, or varying coefficient terms.

As an example, consider the case-study on forest health that will be presented in full detail in Section 5.2. Here we aim at analyzing the impact of several covariates on the health status of trees measured as a binary indicator at a number of observation plots repeatedly over time. Instead of a linear predictor, previous analyses have suggested a model with a geoadditive predictor

$$\eta = \boldsymbol{x}'\boldsymbol{\beta} + f_1(x_1) + \ldots + f_q(x_q) + f(x_1, x_2) + f_{\text{spatial}}(s_1, s_2) + b_{\text{plot}},$$

where $\boldsymbol{x}'\boldsymbol{\beta}$ contains usual linear effects of, for example, categorical covariates, $f_1(x_1), \ldots, f_q(x_q)$ are smooth functions of continuous covariates such as time or age of the trees, $f(x_1, x_2)$ is an interaction surface, $f_{\text{spatial}}(s_1, s_2)$ is a spatial effect defined upon the Gauß-Krüger coordinate information (s_1, s_2) and b_{plot}

is a plot-specific random effect.

Variable selection and model choice questions arising in such geoadditive models are as follows: Should a continuous covariate be included into the model at all and if so as a linear effect or as a nonparametric, flexible effect? Is the spatial effect required in the model, i.e., is spatial correlation present beyond the spatial variation accounted for by spatially varying covariates? Is the interaction effect required in the model? To answer these questions, we propose a systematic, fully automated approach to model choice and variable selection in geoadditive regression models utilizing a componentwise boosting procedure. Our approach generalizes previous suggestions for generalized additive models by Bühlmann and Yu (2003) and Tutz and Binder (2006) to geoadditive models including space-varying and random effects.

After introducing extended geoadditive regression models and componentwise boosting in Section 2, we propose suitable base-learners for a variety of modelling strategies in Section 3. The main ingredient are penalized splines and their bivariate tensor product extensions. One major difficulty is to obtain base-learners that are comparable in complexity to avoid biased selection towards more flexible effects. The equivalent degrees of freedom of a nonparametric effect will be used as a general measure of complexity for the base-learners and a suitable re-parametrization will allow us to specify any desired degree of freedom for a base-learner.

To demonstrate the flexibility of the presented approach and the variety of model choice problems that can be accomplished with it, we present two case studies in Section 5. In the first example, an analysis of habitat suitability based on species abundance data, we will demonstrate the impact of spatial

correlation on variable selection as well as model choice in geoadditive models and models with space-varying coefficients. In the second example, forest health data are analyzed based on a complex model including nonparametric, spatial, interaction and random effects.

2. Generic Model Representation

Suppose that observations $(y_i, \boldsymbol{z}_i)$, $i = 1, \ldots, n$, have been observed on a response variable y_i and a covariate vector $\boldsymbol{z}_i$ comprising different types of covariates. The conditional expectation of y is related to the covariates in a GLM-type manner via $\mathbb{E}(y|\boldsymbol{z}) = h(\eta(\boldsymbol{z}))$, where h is the fixed inverse link function. However, in contrast to GLMs the function $\eta(\boldsymbol{z})$ is no longer restricted to a linear function of the covariates but replaced by an additive function of r components

$$\eta(\boldsymbol{z}) = \beta_0 + \sum_{j=1}^{r} f_j(\boldsymbol{z}). \tag{1}$$

The functions f_j define generic representations of different types of covariate effects, similar as in structured additive regression models considered in Fahrmeir et al. (2004). To make the model formulation more concrete, consider the following examples of functions f_j: (i) *Linear components* $f(\boldsymbol{z}) = f_{\text{linear}}(x) = x\beta$, where x is a univariate component of the vector $\boldsymbol{z}$ and β is the corresponding regression coefficient. (ii) *Nonparametric, smooth components* $f(\boldsymbol{z}) = f_{\text{smooth}}(x)$, where x is a continuous component of $\boldsymbol{z}$ and f_{smooth} is a function of x satisfying certain smoothness conditions. (iii) *Spatial effects and interaction surfaces* $f(\boldsymbol{z}) = f_{\text{spatial}}(x_1, x_2)$, where x_1 and x_2 are continuous covariates and f_{spatial} is a smooth, bivariate surface. In case of a spatial effect, x_1 and x_2 represent coordinate information on the spatial location where an

observation has been collected. (iv) *Varying coefficient terms* (Hastie and Tibshirani, 1993) $f(\boldsymbol{z}) = x_1 f_{\text{smooth}}(x_2)$ or $f(\boldsymbol{z}) = x_1 f_{\text{spatial}}(x_2, x_3)$, where the interaction variable x_1 is either a continuous or a binary covariate, the effect modifiers x_2 (and x_3) are continuous covariates, and $f(\cdot)$ is either a smooth univariate or a smooth bivariate function. If coordinate information is used as effect modifier, the resulting models are also called models with space-varying effects of geographically weighted regression models. (v) *Cluster-specific random effects* $f(\boldsymbol{z}) = b_c$ or $f(\boldsymbol{z}) = x_1 b_c$, where c is a cluster index that relates an observation to the corresponding cluster the observation pertains to. For each group, a separate effect b_c is specified which, under appropriate distributional assumptions, defines either a random intercept or a random slope of covariate x_1.

The generic representation allows for a simplified formulation of complex models in terms of a unifying model description. Moreover, it tremendously facilitates the formulation of a generic componentwise boosting algorithm for variable selection and model choice, where each model component is represented by a corresponding base-learner. In general, boosting can be interpreted as a functional gradient descent method that seeks the solution of the optimization problem

$$\eta^*(\boldsymbol{z}) = \underset{\eta(\boldsymbol{z})}{\operatorname{argmin}}\, \mathbb{E}(\rho(y, \eta(\boldsymbol{z})), \tag{2}$$

where $\rho(\cdot, \cdot)$ is a suitable loss function such as the quadratic (L_2-)loss $\rho(y, \eta) = 0.5|y - \eta|^2$ or the (log-)likelihood function. In practice, (2) is replaced by the empirical risk

$$\frac{1}{n} \sum_{i=1}^{n} \rho(y_i, \eta(\boldsymbol{z}_i))$$

and the boosting algorithm minimizes this quantity with respect to η. After initializing the function estimate with a suitable starting value $\hat{\eta}^{[0]}$, the boosting procedure iteratively computes the negative gradient

$$u_i = -\frac{\partial}{\partial \eta}\rho(y_i, \eta)\Big|_{\eta=\hat{\eta}^{[m-1]}(\boldsymbol{z}_i)}, \; i = 1, \ldots, n$$

evaluated at the current function estimate and fits a base-learner g to $\boldsymbol{u} = (u_1, \ldots, u_n)'$. Since we are not only interested in obtaining an estimate of η but mainly in model choice and variable selection, we utilize a componentwise boosting algorithm. That means, we specify separate base-learners g_j that correspond to the functions f_j which define η. Then, we select the best-fitting componentwise base-learner

$$j^* = \underset{1 \leq j \leq r}{\operatorname{argmin}} \sum_{i=1}^{n} (u_i - g_j(\boldsymbol{z}_i))^2$$

and update the corresponding function estimate $\hat{f}_j$ to

$$\hat{f}_{j^*}^{[m]}(\cdot) = \hat{f}_{j^*}^{[m-1]}(\cdot) + \nu g_{j^*}^{[m]}(\cdot),$$

where $\nu \in (0, 1]$ is a given step size, see Bühlmann and Hothorn (2008) for a detailed derivation and examples. All other effects are kept constant, i.e., $\hat{f}_j^{[m]}(\cdot) = \hat{f}_j^{[m-1]}(\cdot)$ for all $j \neq j^*$. All base-learners considered in Section 3 can be expressed as penalized least squares fits $g_j(\boldsymbol{z}) = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{K})^{-1}\boldsymbol{X}'\boldsymbol{u}$, where $\boldsymbol{X}$ is a suitable design matrix (specifically introduced in Section 3), $\boldsymbol{K}$ is a penalty matrix and λ the corresponding smoothing parameter.

Variable selection and model choice then reduce to stopping the boosting algorithm after an appropriate number of iterations m_{stop}. Within the m_{stop} first iterations some of the base-learners will never have been selected and,

hence, the boosting algorithm provides a means of variable selection. Utilizing competing base-learners implementing different modelling possibilities for the same covariates also addresses the problem of model choice.

3. Base-learner for Geoadditive Regression Models

3.1 *Nonparametric Effects*

To derive appropriate base-learners for nonparametric effects of univariate continuous covariates, we first introduce a suitable nonparametric function estimate in the setting of scatterplot smoothing. Consider the simple model

$$u_i = g(x_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, n, \tag{3}$$

where g is a smooth function of x. A flexible yet parsimonious method to estimate g is to approximate it by a linear combination of B-spline basis functions $B_k^l(x)$ of degree l, i.e.,

$$f(x) = \sum_{k=1}^{K} \beta_k B_j^l(x),$$

where β_k are regression coefficients which scale the basis functions. In principle, such an approach can be interpreted as a large linear model where the evaluations of the basis functions define the design matrix. This leads to the matrix representation of (3) as $\boldsymbol{u} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ and the the regression coefficients could be estimated by least-squares. However, smoothness and form of the resulting function estimate crucially depend on the number of basis functions employed. To overcome this problem, Eilers and Marx (1996) introduced the idea of penalized splines, where a smoothness penalty is added to the least squares criterion when estimating the regression coefficients. A suitable penalty term can be constructed using an approximation to squared

derivatives of $g(x)$ based on differences of the sequence of regression coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$. The d-th order derivative of a B-spline is essentially determined by the d-th order differences leading to the penalty term

$$\lambda P(\boldsymbol{\beta}) = \lambda \sum_{k=d+1}^{K} \Delta_d(\beta_k)$$

where Δ_d denotes the d-th order difference operator, e.g.,

$$\Delta_1(\beta_k) = \beta_k - \beta_{k-1} \qquad \text{or} \qquad \Delta_2(\beta_k) = \beta_k - 2\beta_{k-1} + \beta_{k-2}$$

for first and second order differences, respectively. Estimation of $\boldsymbol{\beta}$ is then based on the penalized least squares (PLS) criterion

$$(\boldsymbol{u} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{u} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda P(\boldsymbol{\beta}), \tag{4}$$

The smoothing parameter $\lambda \geq 0$ controls the flexibility of the function estimate with large values enforcing smooth estimates and small values allowing for high flexibility. Employing a large number of basis functions yields a flexible representation of the nonparametric effect $g(x)$ where the actual degree of smoothness can be adaptively chosen by varying λ.

Seeking the minimizer of the PLS criterion (4) finally yields the base-learner for nonparametric effects. To obtain a compact representation, we rewrite the penalty term as the quadratic form $\lambda P(\boldsymbol{\beta}) = \lambda \boldsymbol{\beta}' \boldsymbol{K} \boldsymbol{\beta}$ where the penalty matrix $\boldsymbol{K}$ is given by $\boldsymbol{K} = \boldsymbol{D}_d' \boldsymbol{D}_d$ and $\boldsymbol{D}_d$ is a d-th order difference matrix of appropriate dimension. Then the penalized least squares estimate of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{K})^{-1} \boldsymbol{X}'\boldsymbol{u}$ and the corresponding base-learner can be represented in terms of the hat or smoother matrix (Hastie and Tibshirani, 1990) $\boldsymbol{S}_\lambda = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{K})^{-1} \boldsymbol{X}'$ as $g(\boldsymbol{z}) = \boldsymbol{S}_\lambda \boldsymbol{u}$.

When performing model choice in semiparametric regression models, a crucial point in defining the nonparametric base-learner is the appropriate choice of the smoothing parameter λ. If we choose λ too large, this will lead to a bias in the boosting selection process preferring nonparametric effects over parametric effects due to their additional flexibility. In addition, we would like to select smoothing parameters that make the nonparametric effects of different covariates comparable in terms of their complexity. A natural measure built in analogy to model complexity in linear models is to consider the trace of the smoother matrix $\boldsymbol{S}_\lambda$ as equivalent degrees of freedom

$$\text{df}(\lambda) = \text{trace}(\boldsymbol{S}_\lambda) = \text{trace}(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \lambda\boldsymbol{K})^{-1}\boldsymbol{X}') = \text{trace}((\boldsymbol{X}'\boldsymbol{X} + \lambda\boldsymbol{K})^{-1}\boldsymbol{X}'\boldsymbol{X}),$$

see Hastie and Tibshirani (1990). Degrees of freedom are a general measure for the complexity of a function estimates that allows to compare the smoothness even for different types of effects (e.g. nonparametric versus spatial effects) and for covariates measured on extremely different scales. If the smoothing parameter is set to zero, $\text{df}(\lambda)$ reduces to the usual complexity measure of a linear model, i.e., the number of parameters describing the spline ($\text{df}(\lambda) = K$). Positive values of λ lead to an effective reduction of the number of parameters ($\text{df}(\lambda) < K$). However, even for very large values of the smoothing parameter, we can not make $\text{df}(\lambda)$ arbitrarily small since a $(d-1)$-th order polynomial in x remains unpenalized by the d-th order difference penalty (provided that the degree of the spline is larger than or equal to the order of the difference penalty). Therefore, for differences $d \geq 2$ we can not achieve $\text{df}(\lambda) = 1$ to make the nonparametric effect comparable in complexity to a single parametric effect. For $d = 1$, $\text{df}(\lambda) = 1$ is obtained

in the limiting case $\lambda \to \infty$ since then the estimated effect is equal to a horizontal line and therefore effectively vanishes.

As a consequence, we have to modify the parametrization of the penalized spline. The aim is to split the function $g(x)$ into a parametric part capturing the $(d-1)$-th order polynomial that remains unpenalized and the deviation from this polynomial $g_{\text{centered}}(x)$, i.e.,

$$g(x) = \beta_0 + \beta_1 x + \ldots + \beta_{d-1} x^{d-1} + g_{\text{centered}}(x). \tag{5}$$

We can then assign the parametric effects describing the polynomial part to the usual linear effects and treat each of them separately using a parametric base-learner. For the deviation part one can choose the smoothing parameter such that it has exactly one degree of freedom despite still being a nonparametric effect. Additionally, this re-parameterization has the advantage that the boosting algorithm provides a possibility to check whether the nonparametric modelling approach is needed, simultaneously with answering the question of whether x has any influence on the response at all. If none of the components in (5) is selected, then x has obviously no effect. If only the parametric components are selected, no nonparametric component is needed and the effect can fully be explained in a simplified model with parametric effects only. We will illustrate this point in the applications in Section 5. Note that decomposition (5) is similar in spirit to the truncated power series basis for polynomial splines but using a B-spline based decomposition retains the advantageous numerical behavior of this basis. Technically, the decomposition of $g(x)$ is achieved by decomposing the vector of regression coefficients $\boldsymbol{\beta}$ into its penalized and its unpenalized component, see Fahrmeir

et al. (2004) for a detailed description in the context of mixed model based estimation of geoadditive regression models.

3.2 *Spatial Effects and Interactions*

For spatial effects based on continuous coordinate information (x_1, x_2) or bivariate interaction surfaces of continuous covariates (x_1, x_2), we extend the concept of penalized spline base-learners to two dimensions. Therefore, we first replace the univariate basis functions by their tensor products, i.e.,

$$g_{\text{spatial}}(x_1, x_2) = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \beta_{k_1,k_2} B_{k_1,k_2}(x_1, x_2)$$

where $B_{k_1,k_2}(x_1, x_2) = B_{k_1}(x_1)B_{k_2}(x_2)$, see also Dierckx (1993) and in particular Wood (2006) where basis functions and products are discussed extensively. In a similar way as for univariate nonparametric effects, this leads to a representation of the vector of spatial effects as the product of a design matrix $\boldsymbol{X}$ containing the evaluations of the tensor product basis functions and a vector of regression coefficients $\boldsymbol{\beta} = (\beta_{11}, \ldots, \beta_{K_1,1}, \ldots, \beta_{1,K_2}, \ldots, \beta_{K_1,K_2})'$, which is the vectorized representation of the bivariate field of regression coefficients. To construct a penalty term in analogy to univariate penalized splines as in Eilers and Marx (2003), we consider separate penalties in x_1 and x_2 direction first. The former can be obtained by constructing a univariate penalty matrix $\boldsymbol{K}_1$ of dimension $(K_1 \times K_1)$ and applying this matrix to each of the subvectors of $\boldsymbol{\beta}$ corresponding to a row in x_1 direction. In matrix notation, this can be facilitated by blowing up $\boldsymbol{K}_1$ based on the Kronecker product with a K_2-dimensional identity matrix, yielding the penalty term $\boldsymbol{\beta}'(\boldsymbol{K}_1 \otimes \boldsymbol{I}_{K_2})\boldsymbol{\beta}$. Similarly, a penalty term in x_2-direction is obtained as $\boldsymbol{\beta}'(\boldsymbol{I}_{K_1} \otimes \boldsymbol{K}_2)\boldsymbol{\beta}$. Note that in the latter expression the univariate penalty matrix $\boldsymbol{K}_2$ has to be pre-

multiplied with the identity matrix due to the ordering of the elements in $\boldsymbol{\beta}$. Summing up both components finally leads to the bivariate penalty term

$$\lambda \boldsymbol{\beta}' \boldsymbol{K} \boldsymbol{\beta} = \lambda \boldsymbol{\beta}' (\boldsymbol{K}_1 \otimes \boldsymbol{I}_{K_2} + \boldsymbol{I}_{K_1} \otimes \boldsymbol{K}_2) \boldsymbol{\beta}$$

which penalizes variation in both x_1 and x_2 direction. A base-learner for spatial and interaction effects is then given by $\boldsymbol{S}_\lambda = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{K})^{-1}\boldsymbol{X}'$ which resembles the base-learner for univariate effects despite the increased number of regression coefficients involved in the description of surfaces.

As for univariate nonparametric smoothing, the spatial effect has to be decomposed into a parametric component representing the unpenalized part of the function estimate and the penalized deviation. From the construction of the penalty term it can be deduced that the unpenalized part is the tensor product of the univariate unpenalized parts. For example, in case of second order differences in both x_1 and x_2 direction, a polynomial of degree one remains unpenalized for both x_1 and x_2. The tensor product of these two linear effects is then represented by an intercept, a linear effect in x_1, a linear effect in x_2 and the interaction $x_1 \cdot x_2$. The deviation effect $g_{\text{centered}}(x_1, x_2)$ can then be constructed in analogy to the univariate setting, see Kneib and Fahrmeir (2006) for details.

3.3 *Varying Coefficient Terms*

Varying coefficient terms (Hastie and Tibshirani, 1993) offer a special way to include interactions between covariates of the form $x_1 f(x_2)$. This can be interpreted as a flexible alternative to a parametric effect $x_1\beta$, where the constant effect β is replaced by a flexible effect function $f(x_2)$. As a special case, varying coefficient models allow to estimate separate effects for

a continuous covariate x_2 in subgroups defined by a binary variable x_1 when employing a predictor of the form

$$\eta(\boldsymbol{z}) = \ldots + f_{\text{smooth},1}(x_2) + x_1 f_{\text{smooth},2}(x_2) + \ldots$$

If $x_1 = 0$, the effect of x_2 is given by $f_{\text{smooth},1}(x_2)$ whereas for $x_1 = 1$, the effect is composed as the sum $f_{\text{smooth},1}(x_2) + f_{\text{smooth},2}(x_2)$ and $f_{\text{smooth},2}(x_2)$ can be interpreted as the deviation effect of x_2 for the group defined by $x_1 = 1$.

Since $f_{\text{smooth},2}(x_2)$ is again a flexible function, we represent the corresponding base-learner in terms of a penalized spline yielding the expression $\text{diag}(x_{11}, \ldots, x_{n1})\boldsymbol{X}^*\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta}$ for the vector of function evaluations $(x_{11}g(x_{12}), \ldots, x_{n1}g(x_{n2}))'$ in matrix notation. The design matrix $\boldsymbol{X}^*$, consisting of the B-spline basis functions representing the varying coefficient, is pre-multiplied by a diagonal matrix containing the values of the interaction variable x_1, yielding the row-wise scaled matrix $\boldsymbol{X}$. The penalty term needs not to be accommodated leading to the base-learner $\boldsymbol{S}_\lambda = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \lambda\boldsymbol{K})^{-1}\boldsymbol{X}'$. In complete analogy we can set up models with space-varying effects $x_1 f_{\text{spatial}}(x_2, x_3)$ where f_{spatial} is a bivariate penalized spline.

To allow for varying coefficient terms with one degree of freedom, restrictions have to be imposed on the base-learner, leading, for example, to

$$x_1 g(x_2) = \beta_0 x_1 + \beta_1 x_1 x_2 + \ldots + \beta_{d-1} x_1 x_2^{d-1} + x_1 g_{\text{centered}}(x_2).$$

3.4 *Random Effects*

For clustered or longitudinal data, correlations between individual observations can be accommodated by the inclusion of random effect terms, leading to a predictor of the form

$$\eta(\boldsymbol{z}) = \ldots + b_{c_i,0} + x_{i1} b_{c_i,1} + \ldots$$

where $c_i \in \{1, \ldots, C\}$ denotes the cluster observation i pertains to. For simplicity, we assume that the clusters are ordered consecutively from 1 to C. In case of longitudinal data, the clusters are defined by individuals whereas the repeated measurements forming the single observations are indexed by i. We utilize the standard assumption of Gaussian random effects, i.e., $b_{c_i,0} \sim \mathcal{N}(0, \tau_0^2)$ is a group-specific random intercept and $b_{c_i,1} \sim \mathcal{N}(0, \tau_1^2)$ is a group-specific random slope.

The corresponding base-learner can then be cast into a similar framework as penalized splines and spatial effects. More specifically, the vector of random intercept evaluations for the observations $i = 1, \ldots, n$ can be expressed as matrix-vector product $\boldsymbol{X}_0 \boldsymbol{b}_0$ where $\boldsymbol{b}_0 = (b_{1,0}, \ldots, b_{C,0})'$ is a vector collecting all random intercepts and $\boldsymbol{X}_0$ is a zero-one incidence matrix that links each observation with the corresponding random intercept. Random slopes can also be considered as varying coefficient terms with a random intercept as effect modifier. For the vector of effects $x_{i1} b_{c_i,1}$ one obtains the expression $\text{diag}(x_{11}, \ldots, x_{n1}) \boldsymbol{X}_0 \boldsymbol{b}_1 = \boldsymbol{X}_1 \boldsymbol{b}_1$. A random effects base-learner is then given by $\boldsymbol{S}_\lambda = \boldsymbol{X}_k (\boldsymbol{X}_k' \boldsymbol{X}_k + \lambda_k \boldsymbol{I}_C) \boldsymbol{X}_k'$, $k = 0, 1$, where λ_k is a smoothing parameter which is inverse proportional to the corresponding random effects variance.

4. Boosting in Geoadditive Regression Models

4.1 *Generic representation*

The generic representation of geoadditive regression models introduced in Section 2 allows for a compact model formulation and description. However, the concept is not limited to model description but can be continued for the formulation of base-learners. As we have seen in Section 3, for all types of

effects in a geoadditive regression model, the base-learners take the form

$$\boldsymbol{S}_\lambda = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{K})^{-1}\boldsymbol{X}'$$

where λ is an appropriately chosen smoothing parameter and $\boldsymbol{K}$ is a penalty matrix. For fixed effects, the smoothing parameter is fixed at zero to obtain an unpenalized fit. For all remaining effects, λ is chosen such that the effect has exactly one degree of freedom, i.e., $\mathrm{df}(\lambda) = \mathrm{trace}(\boldsymbol{S}_\lambda) = 1$. Note, that the degrees of freedom do not depend on the response variable. This is crucial for an efficient implementation of the boosting algorithm, since then the response variable is iteratively replaced by working residuals while proceeding through the fitting process. The desired value for the smoothing parameter can be obtained via a simple line search, although more sophisticated approaches can be used to speed up computations. The search for the smoothness parameter has to be performed only once in a setup step for the algorithm prior to the actual estimation loop.

4.2 *A unified boosting algorithm*

Utilizing the generic representation, the geoadditive regression model has the form

$$\eta(\boldsymbol{z}) = \beta_0 + \sum_{j=1}^{r} f_j(\boldsymbol{z})$$

where $f_j(\boldsymbol{z})$ represent the candidate functions of the predictor. A componentwise boosting procedure based on the loss function $\rho(\cdot)$ can be summarized as follows:

1. Initialize the model components as $\hat{f}_j^{[0]}(\boldsymbol{z}) \equiv 0$, $j = 1, \ldots, r$. Set the iteration index to $m = 0$.

2. Increase m by 1. Compute the current negative gradient

$$u_i = -\frac{\partial}{\partial \eta}\rho(y_i, \eta)\Big|_{\eta=\hat{\eta}^{[m-1]}(\boldsymbol{z}_i)}, \quad i = 1, \ldots, n.$$

3. Choose the base-learner g_{j^*} that minimizes the L_2-loss, i.e. the best-fitting function according to

$$j^* = \underset{1\leq j\leq r}{\operatorname{argmin}} \sum_{i=1}^{n} (u_i - \hat{g}_j(\boldsymbol{z}_i))^2$$

4. Update the corresponding function estimate to

$$\hat{f}_{j^*}^{[m]}(\cdot) = \hat{f}_{j^*}^{[m-1]}(\cdot) + \nu \boldsymbol{S}_{j^*} \boldsymbol{u},$$

where $\nu \in (0, 1]$ is a step size. For all remaining functions set $\hat{f}_j^{[m]}(\cdot) = \hat{f}_j^{[m-1]}(\cdot)$, $j \neq j^*$.

5. Iterate steps 2 to 4 until $m = m_{\text{stop}}$.

Typically, the loss function ρ is given by the log-likelihood of the exponential family under consideration. For the quadratic L_2-loss, the negative gradient equals the working residuals. Note that the boosting procedure used here differs from LogitBoost as considered by Friedman et al. (2000) for binary response and by Tutz and Binder (2006) for exponential family responses, where in each boosting iteration a one step penalized likelihood fit with weights provides the base-learner. In the present approach, a penalized least squares fit to the current negative gradient vector is used instead.

To complete the specification of the boosting algorithm, appropriate values for the step-width ν and for m_{stop} have to be defined. The step-width is typically taken to be relatively small to dampen the effect of the current

fit. We use $\nu = 0.1$ which has proven to be an appropriate default choice. Selection of m_{stop} can be based on AIC reduction: As long as a further iteration decreases AIC, increase the iteration index until a minimum is reached. To avoid local minima, it is typically favorable to fit a larger number of iterations, trace the evolution of AIC with the iteration index and to use the minimum as m_{stop} only if it is far enough from the largest iteration that has been fitted. Typically, AIC decreases much faster for the early iterations, whereas the increase after the minimum is much slower. This represents a convenient property of boosting procedures usually termed as slow overfitting behavior. Even if we chose m_{stop} considerably larger than the optimal value, the resulting model would typically still fit the data reasonable well — see Bühlmann and Hothorn (2008) for an explanation and Bühlmann (2006) for the derivation of AIC based on the output of a boosting algorithm.

The basic difficulty with an AIC-based selection of m_{stop} is the requirement for evaluating the hat matrix defining the best-fitting base procedure in every iteration. Since the hat matrix is of dimension $n \times n$, these computations will be slow if not infeasible for larger data sets. In such cases, bootstrapping is an alternative strategy to determine m_{stop}.

5. Applications

5.1 *Habitat Suitability for Breeding Bird Communities*

In our first application, we analyze counts of subjects from breeding bird communities collected at 258 observation plots in the "Northern Steigerwald", a forest area of about 10.000 hectare, located in northern Bavaria (Müller, 2005). The major aim of this study is to identify factors influencing

habitat suitability and we will employ geoadditive extensions of log-linear Poisson GLMs to accomplish this task.

Originally, 43 species of diurnal breeding birds were sampled five times at each observation site from March to June 2002 using a quantitative grid mapping. To obtain conclusions regarding habitat quality that are more robust and universally valid, species having similar habitat requirements are collected in seven structural guilds (SG) as defined in Table 1. For each site, 31 habitat factors (see Table 2) were measured, describing different aspects of the habitat selection process.

Variable Selection in GLMs with Spatial Component In a first step, we investigated the impact of spatial correlation on variable selection properties in generalized linear models. Therefore we fitted log-linear Poisson regression models with the 31 influential variables from Table 2 entering in linear form. Besides the purely linear model ignoring spatial correlation, we considered spatial models including a bivariate penalized spline surface of the coordinates. We utilized a first order difference penalty and 12 inner knots for each of the directions. In a first spatial GLM approach, five degrees of freedom were assigned to the spatial base-learner, which allows for considerably more flexibility of the spatial effect compared to the remaining linear effects in the regression model. To investigate the impact of this positive discrimination, we considered a spatial GLM where the spatial base-learner is centered as described in Section 3 and can therefore be assigned exactly one degree of freedom making it comparable to a parametric effect.

Table 3 shows the relative selection frequencies for structural guild SG4

obtained from the three models. Comparing the non-spatial GLM and the high degree of freedom spatial GLM first, the inclusion of the spatial effect has a tremendous effect on variable selection, in particular reducing the inclusion frequencies for several of the covariates, such as DWC. Reducing the degrees of freedom to one considerably changes the picture. Now most of the selection frequencies are relatively close to the corresponding value from the non-spatial GLM, although some reduced frequencies still reflect the influence of spatial correlations on the selection process. Also, the selection frequency for the spatial effect itself is largely reduced when reducing the degrees of freedom. This, in turn, shows up in the resulting spatial effect visualized in Figure 1: Both for high and low degrees of freedom, the spatial effect follows essentially the same pattern but is considerably lowered in the latter case. A qualitatively similar behavior is found for the other structural guilds as well, although in some cases, where the spatial effect is not very expressed, the inclusion frequency may even be increased in models with one degree of freedom.

Geoadditive Models In a next step, we extended the spatial GLM to a geoadditive model, where all covariates are allowed for possibly non-linear effects (except for LCA which has only 5 distinct values and is therefore not suitable for nonparametric modelling). All nonparametric effect base-learners are specified as penalized splines with second order difference penalty and 20 inner knots for the spline basis. The spatial base-learner is again included as a bivariate penalized spline with first order difference penalty and 12 inner knots for each coordinate. To differentiate between a flexible nonparametric

effect, parametric linear effect and no effect of a covariate, all nonparametric base-learners were centered around their unpenalized component, i.e. the linear part is effectively substracted from the nonparametric effect. Consequently, linear base-learners of all covariates are included separately into the boosting algorithm.

For structural guild 5, 24 out of the 31 possible covariates were identified to have at least some impact on habitat suitability. Three out of them (DIO, GAP, AGR) only appeared as linear effects in the selected model, whereas the remaining 21 (GST, DBH, AOT, AFS, DWC, LOG, COO, CRS, HRS, OAK, COT, ALA, MAT, ROA, HOT, CTR, BOL, MSP, MDT, MAD, COL) appeared either as purely nonparametric or as the sum of a linear and a nonparametric component. Some selected effects (corresponding to the variables selected most frequently) are visualized in Figure 3 and the spatial effects estimated in the model is displayed in Figure 2. Nonlinear modelling of covariate effects also allows for deeper insight into the habitat selection process of the species. In stands with very low and very high DBH, gaps in the canopy result in higher abundance. A similar interpretation holds for COO where the effect is relatively flat over a wide range, corresponding to the fact that beeches do not need too much light for regeneration. For AOT, 100 years is the age of trees where most felling operations are observed. This results in gaps and following regeneration which leads to a higher abundance.

Space-Varying Effects Finally, we investigated whether some of the covariate effects are spatially-varying and therefore considered varying coefficient models where the continuous covariates enter as interaction variables

in a model with a bivariate surface of the coordinates as effect modifier. For all spatial base-learners, first order differences and 12 inner knots where applied and a purely spatial effect without interaction variable was included in addition. All spatial base-learners are centered, allowing to assign one degree of freedom to each of them. The covariates where additionally included as linear effects, allowing to discriminate between the absence of any effect, a linear effect and a non-linear space-varying effect. For guild 3, 13 variables (GST, AFS, LOG, COM, OAK, ALA, MAT, GAP, AGR, LCA, SCA, MAD, AGL) had exclusively linear influence on habitat suitability. For 12 further variables (DWC, CRS, PIO, GAP, AGR, ROA, SCA, HOT, BOL, MSP, MDT, SUL), spatially varying effects were identified, some of which are shown in Figure 4. Interestingly, the spatial effect without interaction variable was never selected, indicating that all spatial correlation is in fact covered by space-varying effects of some of the covariates. The effects for DWC and ROA correspond to a higher abundance of dead wood and road density, respectively, which also seems to modify the corresponding effect. Similarly, for BOL patchiness is higher in the north east due to small scale cutting resulting in increased heterogeneity and therefore longer borderlines.

5.2 *Forest Health*

The data set considered in the second application is more complex than the first example: The health status of beeches at 83 observation plots located in a northern Bavarian forest district has been assessed in visual forest health inventories carried out between 1983 and 2004. Originally, the health status is classified on an ordinal scale, where the nine possible categories denote

different degrees of defoliation. The domain is divided in 12.5% steps, ranging from healthy trees (0% defoliation) to trees with 100% defoliation. Since data become relatively sparse already for a medium amount of defoliation, we will model the dichotomized response variable defoliation with categories 1 (defoliation above 25%) and 0 (defoliation less or equal to 25%). Table 4 contains a brief description of the covariates in the data set.

Obviously, the collected data have both a temporal and a spatial component that has to be considered in the analysis. Moreover, due to the longitudinal structure of the data, we are interested in estimating plot-specific random effects. Previous studies described in Kneib and Fahrmeir (2006) and Kneib and Fahrmeir (2008) also suggest the presence of interaction effects and non-linear influences of some continuous covariates. Based on these results we consider a logit model with candidate predictor

$$\begin{aligned}\eta(\boldsymbol{z}) &= \boldsymbol{x}'\boldsymbol{\beta} + f_1(\text{ph}) + f_2(\text{canopy}) + f_3(\text{soil}) + f_4(\text{inclination}) \\ &\quad + f_5(\text{elevation}) + f_6(\text{time}) + f_7(\text{age}) + f_8(\text{time}, \text{age}) \\ &\quad + f_9(s_1, s_2) + b_{\text{plot}},\end{aligned}$$

where $\boldsymbol{x}$ contains the parametric effects of the categorical covariates and the base-learners for the smooth effects $f_1, \ldots, f_7$ are specified as univariate cubic penalized splines with 20 inner knots and second order difference penalty. For both the interaction effect f_8 and the spatial effect f_9 we assumed bivariate cubic penalized splines with first order difference penalties and 12 inner knots for each of the directions. Finally, the plot-specific random effect b_{plot} is assumed to be Gaussian with random effects variance fixed such that the base-learner has one degree of freedom. Similarly, all univariate and

bivariate nonparametric effects are decomposed into parametric parts and nonparametric parts with one degree of freedom each. Since the number of observations is too large for AIC-based choice of the stopping rule, m_{stop} was determined by a bootstrapping procedure.

After applying the stopping rule, no effect was found for the ph-value, inclination of slope and elevation above sea level. The univariate effects for age and calendar time where strictly parametric linear but the interaction effect turned out to be very influential. The sum of both linear main effects and nonparametric interaction is shown in Figure 5. The spatial effect was selected only in a relatively small number of iterations whereas the random effect was the component selected most frequently. We can therefore conclude that spatial variation in the data set seems to be present mostly very locally, which is also confirmed by the results found in Kneib and Fahrmeir (2008). For canopy density and soil depth nonlinear effects where identified as visualized in Figure 5. In summary, our results resemble those found in previous analyses but have the advantage that model choice and variable selection can be addressed simultaneously with model fitting.

6. Summary

Based on boosting techniques, an approach is presented allowing for variable selection and model choice in rather complex predictor settings in geoadditive modelling. Since purely nonparametric estimation without structuring assumptions is hopeless, our approach starts with a set of candidate terms within the predictor which has a general additive form that includes parametric as well as nonparametric effect structures. The nonparametric part

can be composed of arbitrary combinations of smooth, spatial, interaction, space-varying or random effects. To avoid selection bias towards nonparametric effects, a reparameterisation is introduced that allows to assign exactly one degree of freedom to all effects. The pre-selection of candidate sets can hardly be avoided but is not very restrictive in practical circumstances since it can be chosen very general when using boosting techniques. The proposed procedure automatically simplifies the pre-specified structuring.

All analyses are performed using an R (R Development Core Team, 2007) implementation of the presented methodology available to the interested reader from package **mboost** (Hothorn et al., 2007). Its user-interface (implemented in function `gamboost()`) facilitates the generic representation of geoadditive models introduced in Section 2. Thus, the transition from theory to practice is leveraged by this common modelling language for geoadditive models.

References

Bühlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics* **34**, 559–583.

Bühlmann, P. and Hothorn, T. (2008). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science* (accepted).

Bühlmann, P. and Yu, B. (2003). Boosting with L_2 loss: Regression and classification. *Journal of the American Statistical Association* **98**, 324–338.

Dierckx, P. (1993). *Curve and surface fitting with splines*. New York: Oxford University Press.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing using B-splines and penalties. *Statistical Science* **11**, 89–121.

Eilers, P. H. C. and Marx, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems* **66**, 159–174.

Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized structured additive regression: A Bayesian perspective. *Statistica Sinica* **14**, 731–761.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalize likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

Friedman, J. H., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion). *Annals of Statistics* **28**, 337–407.

George, E. I. (2000). The variable selection problem. *Journal of the American Statistical Association* **95**, 1304–1308.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Boca Raton, Florida: Chapman and Hall.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* **55**, 757–796.

Hothorn, T., Bühlmann, P., Kneib, T., and Schmid, M. (2007). *mboost: Model-Based Boosting*. R package version 0.6-2.
URL http://R-forge.R-project.org

Kneib, T. and Fahrmeir, L. (2006). Structured additive regression for categorical space-time data: A mixed model approach. *Biometrics* **62**, 109–118.

Kneib, T. and Fahrmeir, L. (2008). A space-time study on forest health. In R. Chandler and M. Scott, editors, *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. New York: John Wiley & Sons.

Müller, J. (2005). Forest structures as key factor for beetle and bird communities in beech forests. PhD thesis.
URL http://mediatum.ub.tum.de

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL http://www.R-project.org

Tutz, G. and Binder, H. (2006). Generalized additive modelling with implicit variable selection by likelihood based boosting. *Biometrics* **62**, 961–971.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall / CRC, Boca Raton.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B* **67**, 301–320.

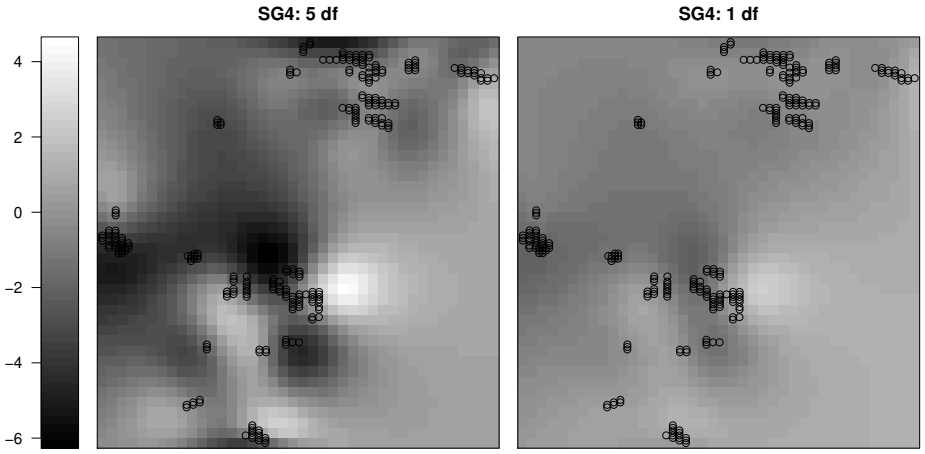

Figure 1. *Guild 4: Estimated spatial effect in GLMs with with either high degrees of freedom or one degree of freedom for the spatial component.*

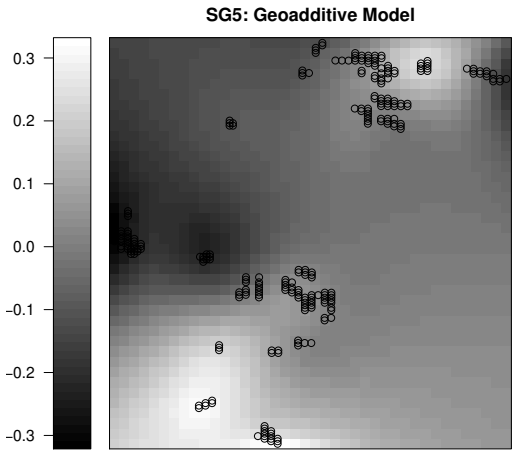

Figure 2. *Guild 5: Estimated spatial effect in a geoadditive model.*

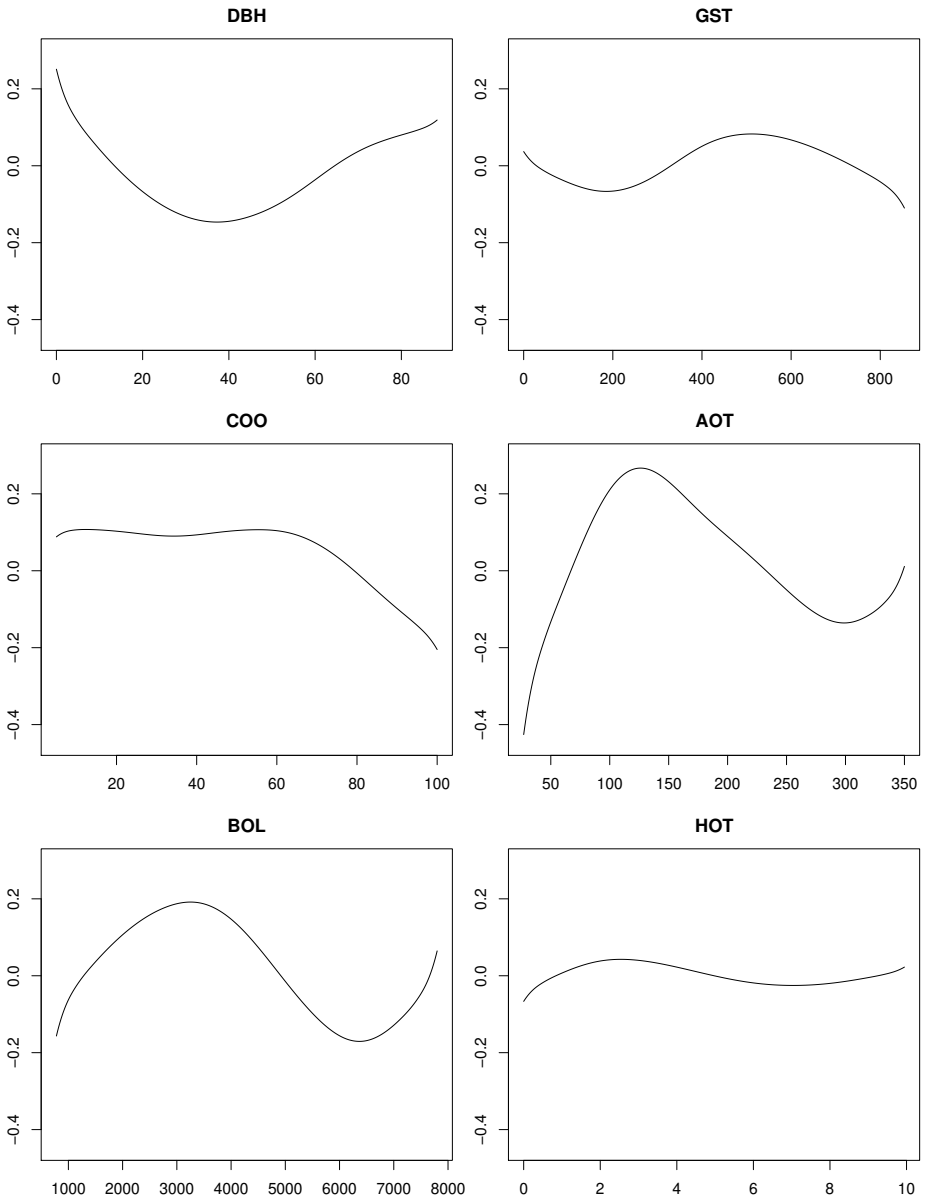

Figure 3. *Guild 5: Selected nonparametric effects in a geoadditive model.*

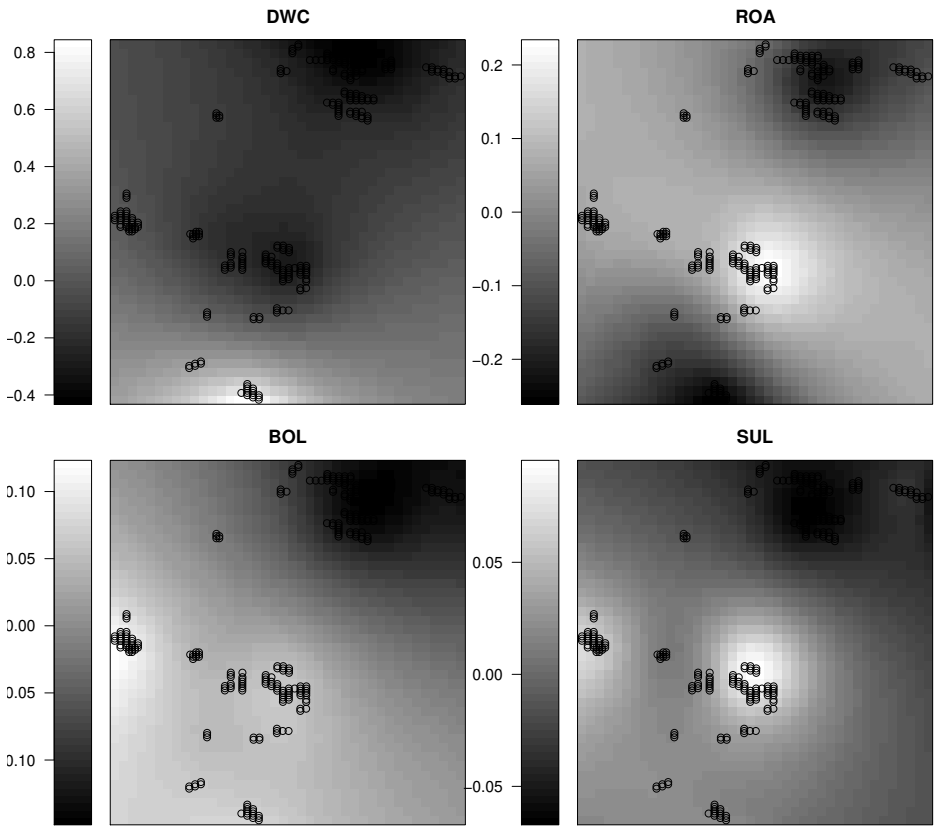

Figure 4. *Guild 3: Selected space-varying effects.*

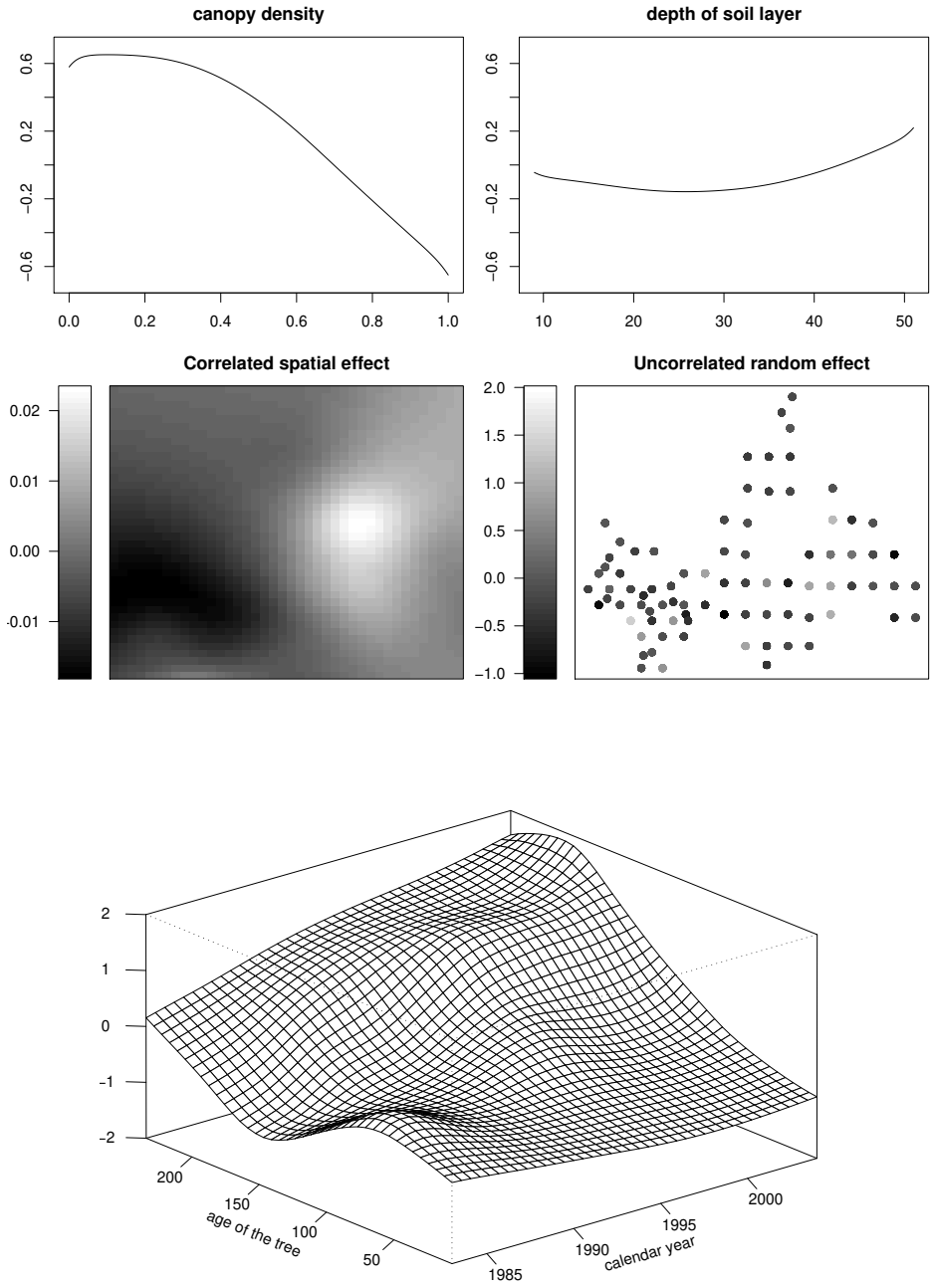

Figure 5. *Forest health: Estimation results.*

Table 1

Definition of structural guilds.

Name	Description	Species
SG1	Requirement of small caves, snags and habitat trees	Ficedula albicollis, F. hypoleuca, F. parva
SG2	Requirement of old beech forests	Dendrocopos medius, D. minor
SG3	Requirement of mature deciduous trees	Sitta europaea, Dendrocopos major, Parus caeruleus, Certhia familiaris
SG4	Requirement of regeneration	Phylloscopus trochilus, Aegithalos caudatus
SG5	Requirement of regeneration combined with planted conifers	Phylloscopus collybita, Turdus merula, Sylvia atricapilla
SG6	Requirement of coniferous trees	Regulus ignicapillus, Parus ater, Prunella modularis
SG7	Requirement of coniferous stands	Regulus regulus, Parus cristatus

Table 2

Environmental variables: Abbreviation, description, range, source and inventory area.

	Description	Range	Source	Inventory
Variables at stand scale				
GST	Growing stock per grid	0-854m^3/ha	Forest inventory	0.05 ha
DBH	Mean diameter of the largest three trees	0-88cm	Forest inventory	0.05 ha
AOT	Age of oldest tree	27-350y	Forest inventory	0.5 ha circle
AFS	Age of forest stand	27-300y	Forest inventory	stand level
DWC	Amount of dead wood of conifers	0-127m^3/ha	Additional inventory	0.1 ha circle
LOG	Amount of logs per grid	0-293m^3/ha	Additional inventory	0.1 ha circle
SNA	Amount of snags and attached dead wood at living trees per grid	0-292m^3/ha	Additional inventory	0.1 ha circle
COO	Canopy over overstorey	5-100%	Estimation in field	1 ha grid
COM	Canopy over middlestorey	0-60%	Estimation in field	1 ha grid
CRS	Percentage of cover of regeneration and shrubs	0-95%	Estimation in field	1 ha grid
HRS	Mean height of regeneration and shrubs	0-10m	Estimation in field	1 ha grid
OAK	Percentage of oak trees	0-40%	Estimation in field	1 ha grid
COT	Percentage of coniferous trees	0-80%	Aerial photo	1 ha grid
PIO	Percentage of pioneer trees (Salix, Betula, Populus)	0-75%	Estimation in field	1 ha grid
ALA	Percentage of alder and ash trees	0-60%	Estimation in field	1 ha grid
MAT	Percentage of cover of mature trees	0-100%	Aerial photo	1 ha grid
GAP	Percentage of gaps per grid	0-19%	Aerial photo	1 ha grid
AGR	Percentage of agricultural land per grid	0-21%	Aerial photo	1 ha grid
ROA	Percentage of roads per grid	0-13%	Aerial photo	1 ha grid
LCA	Number of large cavities per grid	0-15n/ha	Additional inventory	0.5 ha circle
SCA	Number of small cavities per grid	0-33n/ha	Additional inventory	0.5 ha circle
HOT	Hollow trees per grid	0-10n/ha	Additional inventory	0.5 ha circle
CTR	Number of cavity trees per ha	0-14n/ha	Additional inventory	0.5 ha circle
Variables at landscape scale				
RLL	Length of roads at the landscape level	992-12647m	Aerial photo	78.5 ha circle
BOL	Length of patch borderlines	780-7800	Aerial photo	78.5 ha circle
MSP	Mean size of habitat patch	39268-261786	Aerial photo	78.5 ha circle
MDT	Percentage of mature deciduous trees at the landscape level	19-97%	Aerial photo	78.5 ha circle
MAD	Percentage of medium aged deciduous trees at the landscape level	0-69%	Aerial photo	78.5 ha circle
COL	Percentage of coniferous trees at the landscape level	0-77%	Aerial photo	78.5 ha circle
AGL	Percentage of agricultural land at the landscape level	0-41%	Aerial photo	78.5 ha circle
SUL	Percentage of succession at the landscape level	0-24%	Aerial photo	78.5 ha circle

Table 3

Guild 4: Relative selection frequencies of covariates in a non-spatial GLM, a spatial GLM with high degrees of freedom for the spatial component, and a spatial GLM with one degree of freedom for the spatial component.

	GST	DBH	AOT	AFS	DWC	LOG	SNA	COO
non-spatial GLM	0	0	0	0.06	0.3	0	0.01	0
spatial with 5 df	0	0.02	0	0.01	0.05	0	0.01	0
spatial with 1 df	0	0	0	0.06	0.15	0	0	0
	COM	CRS	HRS	OAK	COT	PIO	ALA	MAT
non-spatial GLM	0.03	0.04	0.03	0.05	0.06	0	0.04	0.06
spatial with 5 df	0	0.01	0	0	0	0	0.01	0.05
spatial with 1 df	0.03	0.02	0.02	0.04	0.05	0	0.03	0.04
	GAP	AGR	ROA	LCA	SCA	HOT	CTR	RLL
non-spatial GLM	0.03	0	0	0.1	0.07	0	0	0
spatial with 5 df	0.01	0	0.01	0.01	0.01	0	0	0
spatial with 1 df	0.03	0	0	0.07	0.06	0	0	0
	BOL	MSP	MDT	MAD	COL	AGL	SUL	spatial
non-spatial GLM	0	0.06	0	0	0.05	0	0	0
spatial with 5 df	0	0	0	0	0.03	0	0	0.76
spatial with 1 df	0	0.04	0	0	0.04	0	0	0.3

Table 4

Forest health data: Description of covariates.

Covariate	Description
age	age of the tree in years (continuous, 7≤age≤234)
time	calendar time (continuous, 1983≤time≤2004)
elevation	elevation above sea level in meters (continuous, 250≤elevation≤480)
inclination	inclination of slope in percent (continuous, 0≤inclination≤46)
soil	depth of soil layer in centimeters (continuous, 9≤soil≤51)
ph	ph-value in 0-2cm depth (continuous, 3.28≤ph≤6.05)
canopy	density of forest canopy in percent (continuous, 0≤canopy≤1)
stand	type of stand (categorical, 1=deciduous forest, -1=mixed forest).
fertilisation	fertilisation (categorical, 1=yes, -1=no).
humus	thickness of humus layer in 5 categories (ordinal, higher categories represent higher proportions).
moisture	level of soil moisture (categorical, 1=moderately dry, 2=moderately moist, 3=moist or temporary wet).
saturation	base saturation (ordinal, higher categories indicate higher base saturation).