Leitenstorfer, Tutz:

Generalized Monotonic Regression Based on B-Splines with an Application to Air Pollution Data

Generalized Monotonic Regression Based on B-Splines with an Application to Air Pollution Data

Florian Leitenstorfer, Gerhard Tutz
Ludwig-Maximilians-Universität München
Akademiestraße 1, 80799 München
{leiten,tutz}@stat.uni-muenchen.de

30th May 2005

Abstract

In many studies where it is known that one or more of the certain covariates have monotonic effect on the response variable, common fitting methods for generalized additive models (GAM) may be affected by a sparse design and often generate implausible results. A fitting procedure is proposed that incorporates the monotonicity assumptions on one or more smooth components within a GAM framework. The flexible likelihood based boosting algorithm uses the monotonicity restriction for B-spline coefficients and provides componentwise selection of smooth components. Stopping criteria and approximate pointwise confidence bands are derived. The method is applied to data from a study conducted in the metropolitan area of São Paulo, Brazil, where the influence of several air pollutants like SO_2 on respiratory mortality of children is investigated.

Keywords: Monotonic regression, Generalized additive models, Likelihood based boosting, Air pollution data

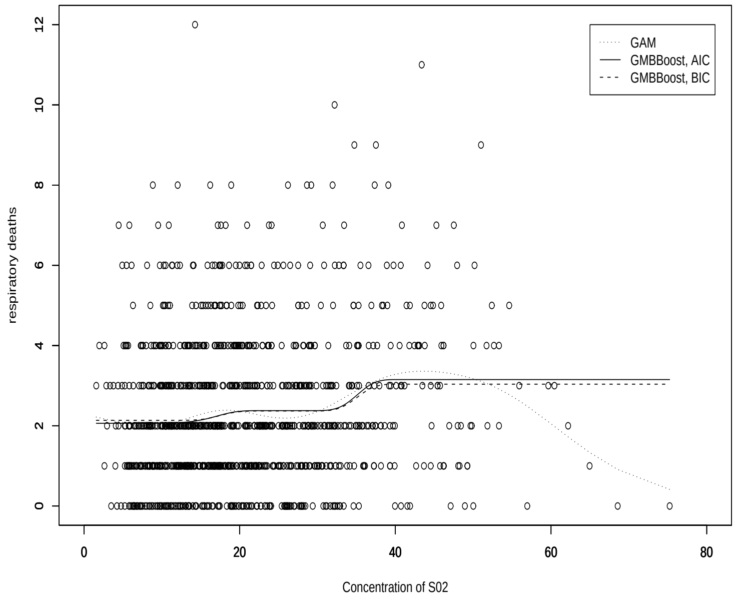
1 Introduction

In many biometrical problems where generalized smooth regression methods are used, a monotonic relationship between one or more explanatory variables and the response variable is to be assumed. A typical problem of this type which will be considered more closely arises in studies where the influence of air pollution on mortality or illness is investigated, see e.g. Schwartz (1994) or Conceição, Miraglia, Kishi, Saldiva & Singer (2001). In these analyses, an increase of deaths or cases of illness is expected with an increasing concentration of a certain pollutant. When standard smoothing techniques, like spline smoothing (Green & Silverman 1994) or local polynomial fitting (Fan & Gijbels 1996), are applied to data of this type in a generalized additive modeling approach, the fitted curves are often affected by few data points. This may lead to unconvincing results. In the following, it is proposed to incorporate the knowledge about monotonic relationships in the estimation by using monotonic regression methods.

Starting from the Pool Adjacent Violators Algorithm (PAVA) (see e.g. Robertson, Wright & Dykstra 1988) which produces a step function, a variety of methods has been developed to smooth the PAVA results, obtaining a smooth estimate of the underlying monotonic function. Details of such approaches, which are mainly based on kernel regression techniques, are given in Friedman & Tibshirani (1984), Mukerjee (1988) or Mammen, Marron, Turlach & Wand (2001). Alternative approaches, which will be pursued in the following, are based on the expansion of a monotonic function into a sum of basis functions, i.e. $f = \sum_j \alpha_j B_j$. To assure monotonicity of the estimate, adequate constraints have to be put on the coefficients α_j. Ramsay (1988) suggests the use of monotonic basis functions (integrated splines), while Kelly & Rice (1991) propose a B-spline basis. As the B-spline approach has become very popular in nonparametric regression (see Eilers & Marx 1996), we will focus on the latter.

Most of the publications on monotonic regression are limited to unidimensional smoothing problems with a Gaussian response variable y. In the example considered here, as in many ecological or biometrical applications, one has multiple covariates $\mathbf{x}' = (x_1, \ldots, x_p)$, and only for some of the covariates a monotonic relationship to $E(y|\mathbf{x})$ has to be assumed. Furthermore, the response variables are typically binary or count data, which are considered as binomial or Poisson distributed. Because little work has been done on monotonic regression in a generalized linear model (GLM) context, least squares approaches have often been used in such cases (see e.g. Kelly & Rice 1991), which lead to dissatisfactory results. Flexible modeling tools are needed, where monotonicity restrictions can easily be incorporated into a generalized additive model (GAM) framework.

Recently, boosting approaches became increasingly important in nonparametric regression, see e.g. Bühlmann & Yu (2003). As Tutz & Leitenstorfer (2005) demonstrate, monotonicity restrictions are easy to include in likelihood based algorithms for generalized response problems by componentwise boosting of monotonic basis functions in each step. In the present paper we suggest boosting based on B-spline basis functions, rather than using monotonic basis functions as in Tutz & Leitenstorfer (2005) or Ramsay (1988). When using B-splines, the monotonicity condition of the estimate is preserved in a different way. A special update scheme for the basis coefficients is proposed which shows good performance. It should be noted that the proposed method avoids the use of algorithms which handle inequality constraints. Procedures of this type typically imply heavy computational burden and often yield unstable estimates. From a Bayesian perspective, a B-spline approach to monotonic regression has been suggested by Brezger & Steiner (2004).



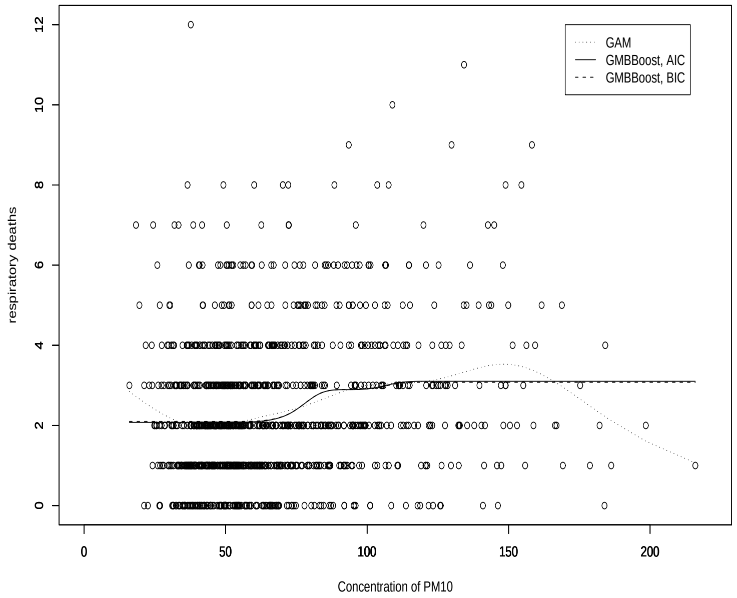


FIGURE 1: *Deaths due to respiratory causes vs. SO_2-concentration (above), and PM_{10}-concentration (below), monotonic Boosting methods and GAM.*

We illustrate generalized monotonic regression techniques on a data set that has been previously analyzed by Conceição, Miraglia, Kishi, Saldiva & Singer (2001), Singer, Andre, Lima & Conçeicão (2002) and Einbeck, Andre & Singer (2004). The data have been collected to evaluate the association between mortality of children under five due to respiratory causes and the concentration of various air pollutants in the city of São Paulo, Brazil, form 1994 to 1997 (the data are available at http://www.ime.usp.br/∼jmsinger; a detailed description follows in Section 4). In Figure 1, in each panel the number of daily respiratory deaths is given as a function of a specific air pollutant, fitted by the proposed monotonic methods and a unconstrained generalized smoothing method (function `gam()` of the R library `mgcv`). The response variable was assumed to be Poisson distributed, and the log-link was used. These examples show that especially for SO_2 and PM_{10}, the GAM fit is pulled downwards by few observations on the right side where the design is sparse. The fitted curves imply a decrease of the mortality for high pollutant concentrations, which has no causal plausibility. In contrast, the monotonic approach shows resistance against such inconsistencies and yields reliable fits. Einbeck, Andre & Singer (2004) found the same problems with the data and proposed to stabilize a local fitting procedure by downweighting points with small design density.

In Section 2 the concept of monotonic likelihood boosting based on B-splines is introduced, and an extension to multiple covariate settings is given. In Section 3 approximate pointwise confidence bands are derived. In section 4 we take a closer look on the data set mentioned above. Note that throughout the paper, we refer to monotonic regression as nondecreasing regression.

2 Boosting B-splines in generalized monotonic regression

2.1 Monotonicity constraints for B-splines

First, we consider a generalized smooth monotonic regression problem with dependent variable y that can be non-Gaussian, and a single covariate x. As in generalized linear models (e.g. McCullagh & Nelder 1989) it is assumed that $y_i|x_i$ has a distribution from a simple exponential family $f(y_i|x_i) = \exp\{(y_i\theta_i - b(\theta_i))/\phi + c(y_i, \phi)\}$ where θ_i is the canonical parameter and ϕ denotes the dispersion parameter. The link between $\mu_i = E(y_i|x_i)$ and the explanatory variable x_i is determined by $\mu_i = h(\eta_i)$, where h is a given response function which is strictly monotone (the inverse of the link function $g = h^{-1}$), and the predictor $\eta_i = \eta(x_i)$ is a function of x. While in generalized linear models, $\eta(x)$ is assumed to be a linear predictor, here more generally it is assumed that $\eta(x) = f(x)$ is a smooth function that satisfies the monotonicity condition

$$f(x) \geq f(z) \quad \text{if} \quad x > z. \tag{1}$$

Obviously, monotonicity in η transforms into monotonicity in the means.

Due to their flexibility, smoothing methods based on B-splines are a common tool in statistics, see e.g. Eilers & Marx (1996). Such approaches are based on an expansion of f into B-spline basis functions, where a sequence of knots $\{t_j\}$ is placed equidistantly within the range $[x_{min}, x_{max}]$. With $\tilde{m}$ denoting the number of interior knots, one obtains the linear term

$$\eta(x) = \alpha_0 + \sum_{j=1}^{m} \alpha_j B_j(x, q), \tag{2}$$

where q denotes the degree of the B-splines and $m = \tilde{m} + 1 + q$ (the augmented set of knots). An algorithm for the computation of B-splines of degree q is given in De Boor (1978). Monotonicity can be assured in the following way: suppose we have B-splines of degree $q \geq 1$ and let h be the distance between the equally spaced knots. Then the derivative $\eta'(x) = \partial\eta(x)/\partial x$ can be written as

$$\eta'(x) = \sum_j \alpha_j B_j'(x, q) = \frac{1}{h} \sum_j (\alpha_{j+1} - \alpha_j) B_j(x, q-1),$$

for a proof see De Boor (1978). Since $B_j(x, q-1) \geq 0$, it follows from

$$\alpha_{j+1} \geq \alpha_j, \tag{3}$$

that $\eta'(x) \geq 0$ holds. In other words, since (3) is a sufficient condition for the monotonicity of $\eta(x)$, the sequence of coefficients α_j has to be nondecreasing in order to obtain monotonic functions. This property of B-splines has been previously exploited by Kelly & Rice (1991) and Brezger & Steiner (2004) in a monotonic regression setting.

2.2 An outline of the algorithm

Boosting has originally been introduced within the machine learning community (e.g. Schapire 1990) for classification problems. More recently, the approach has been extended to regression modeling with a continuous dependent variable (e.g. Bühlmann & Yu 2003, Bühlmann 2004). The basic idea is to fit a function iteratively by fitting in each stage a ”weak” learner to the current residual. In componentwise boosting as proposed by Bühlmann & Yu (2003), only the contribution of one variable is updated in one step. In contrast to these approaches we propose to update a specific simplification of the predictor which makes it easy to control the monotonicity restriction.

For simplicity, in the following, the degree q of the B-splines is suppressed. In matrix notation, the data are given by $\mathbf{y} = (y_1, \ldots, y_n)'$, $\mathbf{x} = (x_1, \ldots, x_n)'$. Based on the expansion into basis function, the data set may be collected in matrix form $(\mathbf{y}, \mathbf{B})$, where $\mathbf{B} = (B_1(\mathbf{x}), \ldots, B_m(\mathbf{x}))$, $B_j(\mathbf{x}) = (B_j(x_1), \ldots, B_j(x_n))'$.

The residual model that is fitted by weak learners in one iteration step uses a grouping of B-splines. One considers for $r = 1, \ldots, m-1$ the simplified model that has the predictor

$$\eta(x_i) = \alpha_{0(r)} + \alpha_{1(r)} \left(\sum_{j=1}^{r} B_j(x_i) \right) + \alpha_{2(r)} \left(\sum_{j=r+1}^{m} B_j(x_i) \right). \tag{4}$$

When fitting model (4) the monotonicity constraint is easily checked by comparing the estimates $\hat{\alpha}_{1(r)}$ and $\hat{\alpha}_{2(r)}$, since monotonicity follows from $\hat{\alpha}_{2(r)} \geq \hat{\alpha}_{1(r)}$. Given an estimate from previous fitting,

$$\hat{\eta}_{old}(x_i) = \hat{\alpha}_{0,old} + \sum_{j=1}^{m} \hat{\alpha}_{j,old} B_j(x_i)$$

refitting is performed by

$$\begin{aligned} \hat{\eta}_{new}(x_i) &= \hat{\eta}_{old}(x_i) + \hat{\alpha}_{0(r)} + \hat{\alpha}_{1(r)} \left(\sum_{j=1}^{r} B_j(x_i) \right) + \hat{\alpha}_{2(r)} \left(\sum_{j=r+1}^{m} B_j(x_i) \right) \\ &= \hat{\alpha}_{0,old} + \hat{\alpha}_{0(r)} + \sum_{j=1}^{r} (\hat{\alpha}_{j,old} + \hat{\alpha}_{1(r)}) B_j(x_i) + \sum_{j=r+1}^{m} (\hat{\alpha}_{j,old} + \hat{\alpha}_{2(r)}) B_j(x_i). \end{aligned}$$

It is obvious that $\hat{\eta}_{new}$ is monotonic if estimates fulfill $\hat{\alpha}_{2(r)} \geq \hat{\alpha}_{1(r)}$, provided that the previous estimate $\hat{\eta}_{old}$ was monotonic. The grouping of basis functions into $B_1, \ldots, B_r$ and $B_{r+1}, \ldots, B_m$ which are adapted by the amount $\alpha_{1(r)}$ in the first and $\alpha_{2(r)}$ in the second group allows to control monotonicity in a simple way. Fitting a full model with m new parameters would imply much more computational effort and rise problems if the newly fitted model is non-monotonic. Instead, the possible groupings ($r = 1, \ldots, m-1$) are evaluated and in analogy to componentwise boosting the best refit is selected. The grouping of B-splines can be derived as a restricted model in the sense of restricted least squares estimators (RLSE) in linear models, see Theil & Goldberger (1961). In the usual form of a smoothed estimate based on B-splines, model (4) is given by

$$\eta(x_i) = \alpha_{0(r)} + \sum_{j=1}^{r} \alpha_j^{(r)} B_j(x_i) + \sum_{j=r+1}^{m} \alpha_j^{(r)} B_j(x_i) \tag{5}$$

with the constraints $\alpha_1^{(r)} = \cdots = \alpha_r^{(r)} = \alpha_{1(r)}$, $\alpha_{r+1}^{(r)} = \cdots = \alpha_m^{(r)} = \alpha_{2(r)}$. The constraints specify that blocks of r and $m-r$ parameters have to be identical.

Before giving the algorithm, which is based on likelihood based boosting strategies as proposed by Tutz & Binder (2004), the fit of model (4) is embedded into the framework of penalized likelihood estimation. Let

$$\mathbf{R}_{(r)} = \begin{pmatrix} \mathbf{1}_r & \mathbf{0}_r \\ \mathbf{0}_{m-r} & \mathbf{1}_{m-r} \end{pmatrix},$$

with $\mathbf{0}_r$, $\mathbf{1}_r$ denoting the vectors of length r containing 0s and 1s only, then (4) may be represented in matrix form by the linear predictor $\eta(x) = \mathbf{B}_{(r)}\boldsymbol{\alpha}_{(r)}$, where $\mathbf{B}_{(r)} = (\mathbf{1}, \mathbf{BR}_{(r)})$ and $\boldsymbol{\alpha}_{(r)} = (\alpha_{0(r)}, \alpha_{1(r)}, \alpha_{2(r)})'$. It is proposed that in each boosting step, model (4) is estimated by one-step Fisher scoring based on generalized ridge regression (Marx, Eilers & Smith 1992). Common ridge regression maximizes the penalized log-likelihood

$$l_p(\boldsymbol{\alpha}_{(r)}) = \sum_{i=1}^{n} l_i(\boldsymbol{\alpha}_{(r)}) - P(\boldsymbol{\alpha}_{(r)}),$$

where $l_i(\boldsymbol{\alpha}_{(r)}) = l_i(h(\mathbf{B}_{(r)}\boldsymbol{\alpha}_{(r)}))$ is the usual log-likelihood contribution of the ith observation and the $P(\boldsymbol{\alpha}_{(r)}) = (\lambda/2)\boldsymbol{\alpha}'_{(r)}\boldsymbol{\alpha}_{(r)}$ represents the penalty term with ridge parameter λ. However, model (4) is asymmetric in a specific sense. Consider therefore the representation (5) of the restricted problem. If for example $r = 2$, the first constraint $\alpha_1^{(2)} = \alpha_2^{(2)}$ concerns only two parameters, whereas the second constraint $\alpha_3^{(2)} = \cdots = \alpha_m^{(2)}$ concerns $m - 2$ parameters, which for $m = 20$ means 18 parameters are restricted. It seems sensible to adapt the penalty to the complexity of the constraints which are implicitly used. We propose to use the penalty

$$P(\boldsymbol{\alpha}_{(r)}) = \frac{\lambda}{2}(r\alpha_{1(r)} + (m - r)\alpha_{2(r)}), \tag{6}$$

where the parameters are weighted by the number of parameters that are implicitly considered as identical. When using (6) we found much better performance of the estimator than by using the usual ridge constraint $P(\boldsymbol{\alpha}_{(r)}) = (\lambda/2)\boldsymbol{\alpha}'_{(r)}\boldsymbol{\alpha}_{(r)}$. In matrix form one obtains the penalized log-likelihood

$$l_p(\boldsymbol{\alpha}_{(r)}) = \sum_{i=1}^{n} l_i(\boldsymbol{\alpha}_{(r)}) - \frac{\lambda}{2}\boldsymbol{\alpha}'_{(r)}\boldsymbol{\Lambda}\boldsymbol{\alpha}_{(r)},$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}'_{(r)}\mathbf{R}_{(r)} \end{pmatrix} = diag(0, r, m - r), \tag{7}$$

and $\lambda > 0$ represents the ridge parameter. Derivation yields the corresponding penalized score function

$$s_p(\boldsymbol{\alpha}_{(r)}) = \frac{\partial l_p(\boldsymbol{\alpha}_{(r)})}{\partial \boldsymbol{\alpha}_{(r)}} = \mathbf{B}'_{(r)}\mathbf{W}(\boldsymbol{\eta})\mathbf{D}(\boldsymbol{\eta})^{-1}(\mathbf{y} - h(\boldsymbol{\eta})) - \lambda\boldsymbol{\Lambda}\boldsymbol{\alpha}_{(r)}, \tag{8}$$

with $\mathbf{W}(\boldsymbol{\eta}) = \mathbf{D}^2(\boldsymbol{\eta})\boldsymbol{\Sigma}(\boldsymbol{\eta})^{-1}$, $\mathbf{D}(\boldsymbol{\eta}) = diag\{\partial h(\eta_1)/\partial\eta, \ldots, \partial h(\eta_n)/\partial\eta\}$, $\boldsymbol{\Sigma}(\boldsymbol{\eta}) = diag\{\sigma_1^2, \ldots, \sigma_n^2\}$, $\sigma_i^2 = var(y_i)$, all of them evaluated at the current value of η. Note that the intercept term is refitted in each iteration by the corresponding unpenalized one-step estimate. The monotonicity constraint from (3) is incorporated by taking into account only estimates which fulfill $\hat{\alpha}_{2(r)} > \hat{\alpha}_{1(r)}$. It is easily seen that the update scheme given below yields the desired nondecreasing sequences of estimates $\hat{\alpha}_1, \dots, \hat{\alpha}_m$ in each boosting iteration. An outline of the algorithm is as follows.

Monotonic Likelihood Boosting for B-splines

Step 1 (Initialization)

Standardize $\mathbf{y}$ to zero mean, i.e. set $\hat{\alpha}_0 = \bar{y}$, $\hat{\boldsymbol{\alpha}}^{(0)} = (\bar{y}, 0, \dots, 0)'$, $\hat{\boldsymbol{\eta}}^{(0)} = (\bar{y}, \dots, \bar{y})'$ and $\hat{\boldsymbol{\mu}}^{(0)} = (h(\bar{y}), \dots, h(\bar{y}))'$.

Step 2 (Iteration)

For $l = 1, 2, \dots$

1. *Fitting step*
For $r = 1, \dots, m-1$ compute the modified ridge estimate based on one step Fisher scoring,
$$\hat{\boldsymbol{\alpha}}_{(r)} = (\mathbf{B}'_{(r)}\mathbf{W}_l\mathbf{B}_{(r)} + \lambda\boldsymbol{\Lambda})^{-1}\mathbf{B}'_{(r)}\mathbf{W}_l{\mathbf{D}_l}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(l-1)}), \tag{9}$$
where $\hat{\boldsymbol{\alpha}}_{(r)} = (\hat{\alpha}_{0(r)}, \hat{\alpha}_{1(r)}, \hat{\alpha}_{2(r)})'$, $\mathbf{W}_l = \mathbf{W}(\hat{\boldsymbol{\eta}}^{(l-1)})$, $\mathbf{D}_l = \mathbf{D}(\hat{\boldsymbol{\eta}}^{(l-1)})$, and $\hat{\boldsymbol{\mu}}^{(l-1)} = h(\hat{\boldsymbol{\eta}}^{(l-1)})$. Compute the potential update of the linear predictor, $\hat{\boldsymbol{\eta}}_{r,new} = \hat{\boldsymbol{\eta}}^{(l-1)} + \mathbf{B}_{(r)}\hat{\boldsymbol{\alpha}}_{(r)}$. Let $A = \{r : \hat{\alpha}_{1(r)} < \hat{\alpha}_{2(r)}\}$ denote the candidates that fulfill the monotonicity constraint.

2. *Selection step*
Compute the potential update of the linear predictor, $\hat{\boldsymbol{\eta}}_{(r),new} = \hat{\boldsymbol{\eta}}^{(l-1)} + \mathbf{B}_{(r)}\hat{\boldsymbol{\alpha}}_{(r)}$, $r \in 1, \dots, m-1$. Choose $r_l \in A$ such that the deviance is minimized, i.e.
$$r_l = \arg\min_{r \in A} Dev(\hat{\boldsymbol{\eta}}_{(r),new}).$$

3. *Update*
Set
$$\begin{aligned} \hat{\alpha}_0^{(l)} &= \hat{\alpha}_0^{(l-1)} + \hat{\alpha}_{0(r_l)}, \\ \hat{\alpha}_j^{(l)} &= \begin{cases} \hat{\alpha}_j^{(l-1)} + \hat{\alpha}_{1(r_l)} & 1 \le j \le r_l \\ \hat{\alpha}_j^{(l-1)} + \hat{\alpha}_{2(r_l)} & j > r_l, \end{cases} \end{aligned} \tag{10}$$
$$\hat{\boldsymbol{\eta}}^{(l)} = \hat{\boldsymbol{\eta}}^{(l-1)} + \mathbf{B}_{(r_l)}\hat{\boldsymbol{\alpha}}_{(r_l)} \quad \text{and} \quad \hat{\boldsymbol{\mu}}^{(l)} = h(\hat{\boldsymbol{\eta}}^{(l)}).$$

When using boosting techniques, the number of iterations l plays the role of a smoothing parameter. Therefore, in order to prevent overfitting, a stopping criterion is necessary. A quite common measure of the complexity of a smooth regression fit is the hat-matrix. Consequently, Bühlmann & Yu (2003) and Bühlmann (2004) developed a hat-matrix for L_2-boosting with continuous dependent variable. In the case of likelihood boosting, for more general exponential type distributions, the hat-matrix has to be approximated. For integrated splines, Tutz & Leitenstorfer (2005) give an approximation based on first order Taylor expansions, which shows satisfying properties. It is straightforward to derive the hat-matrix for the present case along the lines of Tutz & Leitenstorfer (2005). With $\mathbf{M}_0 = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$ and $\mathbf{M}_l = \mathbf{\Sigma}_l^{1/2}\mathbf{W}_l^{1/2}\mathbf{B}_{(r_l)}(\mathbf{B}_{(r_l)}'\mathbf{W}_l\mathbf{B}_{(r_l)} + \lambda\mathbf{\Lambda})^{-1}\mathbf{B}_{(r_l)}'\mathbf{W}_l^{1/2}\mathbf{\Sigma}_l^{1/2}$, where $\mathbf{W}_l = \mathbf{W}(\hat{\boldsymbol{\eta}}^{(l-1)})$, $l = 1, 2, \ldots$, and $\mathbf{\Sigma}_l = \mathbf{\Sigma}(\hat{\boldsymbol{\eta}}^{(l-1)})$, the approximate hat-matrix is given by

$$\mathbf{H}_l = \mathbf{I} - (\mathbf{I} - \mathbf{M}_0)(\mathbf{I} - \mathbf{M}_1)\cdots(\mathbf{I} - \mathbf{M}_l) = \sum_{j=0}^{l} \mathbf{M}_j \prod_{i=0}^{j-1}(\mathbf{I} - \mathbf{M}_i), \tag{11}$$

with $\hat{\boldsymbol{\mu}}^{(l)} \approx \mathbf{H}_l\mathbf{y}$. By considering $tr(\mathbf{H}_l)$ as the degrees of freedom of the smoother, we investigate the AIC and the BIC criteria as potential stopping criteria,

$$AIC(l) = Dev_l + 2tr(\mathbf{H}_l)$$

and

$$BIC(l) = Dev_l + \log(n)tr(\mathbf{H}_l),$$

where $Dev_l = 2\sum_{i=1}^{n}[l_i(y_i) - l_i(\hat{\eta}_i^{(l)})]$ denotes the deviance of the model in the lth boosting step. The optimal number of boosting iterations is defined by $l_{opt}^{AIC} = \arg\min_l AIC(l)$ or $l_{opt}^{BIC} = \arg\min_l BIC(l)$. Since the BIC (Schwarz 1978) penalizes the complexity of the fit stronger, usually more sparse models result. A more extensive treatment of stopping criteria for boosting algorithms is given in Bühlmann & Yu (2005).

2.3 Extension to generalized additive models

In biometrical or ecological problems, one is usually interested in the effect of several predictor variables, where some of them might have monotonic influence on y, whereas others have not. Additionally, a smooth estimation is not always appropriate for all covariates. In the following we demonstrate that the concept given above can easily be extended to a GAM setting (see e.g. Hastie & Tibshirani 1990 or Marx & Eilers 1998). Let

$$\eta(x) = \alpha_0 + \sum_{s=1}^{p} f_s(x_s), \tag{12}$$

where for part of the unknown smooth functions (say $f_1, \ldots, f_v$, $v \leq s$) monotonicity constraints are assumed to hold. Using the matrix notation from above, we have a design matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$, where $\mathbf{x}_s = (x_{1s}, \ldots, x_{ns})'$. Componentwise expansion into B-spline basis functions leads to the data set $(\mathbf{y}, \mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(p)})$, where $\mathbf{B}^{(j)}$ refers to the jth predictor.

It is essential to distinguish between components that are under monotonicity restrictions and those that are not. For the former, grouping of basis functions is done within each component in the same way as described in (4). For the unconstrained components, we follow Bühlmann & Yu (2003) and Tutz & Binder (2004) and use penalized regression splines (P-splines, cf. Eilers & Marx 1996) as weak learner for the chosen component. Thereby, the second order differences of the B-spline coefficients are penalized. For simplicity, it is assumed that the same number of basis functions m is used for all f_s. The vector of basis coefficients for the whole model is then given by $\boldsymbol{\alpha} = (\alpha_0, \alpha_{1,1}, \ldots, \alpha_{1,m}, \ldots, \alpha_{p,1}, \ldots, \alpha_{p,m})'$. Thus, *step 2 (iteration)* of the algorithm described above is extended as follows:

Step 2 (Iteration):

For $l = 1, 2, \ldots$

1. *Fitting step*

   For $s = 1, \ldots, p$,

   - If $s \in \{1, \ldots, v\}$ (the components under monotonicity constraint), compute the estimates from (9) componentwise for the possible groupings $r = 1, \ldots, m-1$, with
   $$\mathbf{B}^{(s)}_{(r)} = (\mathbf{1}, \mathbf{B}^{(s)}\mathbf{R}_{(r)}). \tag{13}$$
   The set of indices for components s and split points r that satisfy the monotonicity constraint is given by
   $$A_1 = \{(s, r) \in \{1, \ldots, v\} \times \{1, \ldots, (m-1)\} : \hat{\alpha}^{(s)}_{1(r)} < \hat{\alpha}^{(s)}_{2(r)}\}.$$
   - If $s \in \{v+1, \ldots, p\}$ (the components without constraints), compute the one step Fisher scoring estimate of the P-spline including the intercept term,
   $$\hat{\boldsymbol{\alpha}}^{(s)} = (\mathbf{B}^{*(s)\prime}\mathbf{W}_l\mathbf{B}^{*(s)} + \lambda_P \boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2)^{-1}\mathbf{B}^{*(s)\prime}\mathbf{W}_l\mathbf{D}_l{}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(l-1)}), \tag{14}$$
   where
   $$\boldsymbol{\Delta}_2 = \begin{pmatrix} 0 & 1 & -2 & 1 & & & \\ \vdots & & \ddots & \ddots & \ddots & \\ 0 & \ldots & & 1 & -2 & 1 \end{pmatrix}$$
   denotes the matrix representation of the second order differences, and $\mathbf{B}^{*(s)} = (\mathbf{1}, \mathbf{B}^{(s)})$. Since the P-spline fit (14) does not distinguish between split points $r \in \{1, \ldots, m-1\}$, for convenience of notation we set

$r = 0$ and extend the selection set by $A_2 = \{(s, 0), s \in \{v+1, \ldots, p\}\}$, yielding

$$A = A_1 \cup A_2. \tag{15}$$

2. *Selection step*
   Compute the potential update of the linear predictor, which only for the monotonic coefficients $s \leq v$ depends on the split point r. Otherwise, r is set to 0, indicating that $\hat{\boldsymbol{\eta}}^{(s)}_{(0),new}$ is not affected by r. Choose $(s_l, r_l) \in A$ such that the deviance is minimized, i.e.

   $$(s_l, r_l) = \arg \min_{(s,r)\in A} Dev(\hat{\boldsymbol{\eta}}^{(s)}_{(r),new}).$$

3. *Update*
   Besides the intercept, in each iteration only the basis coefficients belonging the chosen component s_l are refitted. That means, if the selected s_l is in $\{1, \ldots, v\}$, then $\hat{\alpha}_0^{(l)}$ and $\hat{\alpha}_{s_l,j}^{(l)}$, $j = 1, \ldots, m$, are updated by the refitting scheme (10). If $s_l > v$, then update

   $$\hat{\alpha}_0^{(l)} = \hat{\alpha}_0^{(l-1)} + \hat{\alpha}_0^{(s_l)} \quad \text{and} \quad \hat{\alpha}_{s_l,j}^{(l)} = \hat{\alpha}_{s_l,j}^{(l-1)} + \hat{\alpha}_j^{(s_l)},$$

   with $\hat{\boldsymbol{\alpha}}^{(s_l)}$ from (14).

By using $\mathbf{B}^{(s)}_{(r)}$ from (13), along with $\mathbf{B}^{*(s)}$ and the penalty matrix $\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2$ for the P-spline estimates, the hat-matrix approximation from (11) and the corresponding AIC and BIC stopping criteria can be extended to the additive setting. In the case of many predictors it might occur that boosting stops before a certain component has been chosen. Thus, the extended approach has the nice effect of doing variable selection for smooth components, similar to the methods proposed by Bühlmann & Yu (2003). This additional strength is important only in data sets with a large number p of covariates, where only some of them are influential.

If a set of covariates $x_{w+1}, \ldots, x_p$, $v \leq w < p$ has to be modeled in a parametric way, i.e. if we have a semiparametric linear predictor

$$\eta(x) = \alpha_0 + \sum_{s=1}^{w} f_s(x_s) + \sum_{u=w+1}^{p} \beta_u x_u$$

(e.g. Speckman 1988), the estimation of the corresponding parameters is easy to incorporate in our proposed algorithm. However, it turns out that fixed effects that are included with one basis function in the selection are rarely chosen, especially for dummy covariates, since they carry much less information compared to metrical variables. Hence, after initialization, we treat parametric covariates like the intercept: unpenalized one step estimates result from (9) and (14) respectively by simply enlarging the matrix $\mathbf{B}^{(s)}_{(r)}$ ($\mathbf{B}^{*(s)}$) by $\mathbf{x}_{w+1}, \ldots, \mathbf{x}_p$ and correcting the penalty matrix $\boldsymbol{\Lambda}$ ($\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2$) appropriately. The corresponding coefficients are then updated in each boosting iteration.

3 Standard deviations

In order to obtain standard deviations for function estimates, we suggest to start from the approximate hat-matrix given in (11). Consider the model from (12), where components $1, \dots, v$ are estimated under the monotonicity constraint and components $v+1, \dots, p$ are not, the linear predictor after l boosting iterations is given by

$$\hat{\boldsymbol{\eta}}^{(l)} = \mathbf{1}_n \hat{\alpha}_0^{(l)} + \sum_{s=1}^{p} \mathbf{B}^{(s)} \hat{\boldsymbol{\alpha}}_s^{(l)}, \tag{16}$$

where $\hat{\boldsymbol{\alpha}}_s^{(l)} = (\hat{\alpha}_{s,1}^{(l)}, \dots, \hat{\alpha}_{s,m}^{(l)})'$ and $\hat{\alpha}_0^{(l)}$ results from updating the intercept in each iteration. Let s_k be the component chosen in the kth boosting iteration, one has from the update step of the extended algorithm,

$$\mathbf{B}^{(s)} \hat{\boldsymbol{\alpha}}_s^{(k)} = \begin{cases} \mathbf{B}^{(s)} \hat{\boldsymbol{\alpha}}_s^{(k-1)}, & s_k \neq s \\ \mathbf{B}^{(s)} \hat{\boldsymbol{\alpha}}_s^{(k-1)} + \mathbf{S}_k(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k-1)}), & s_k = s, \end{cases} \tag{17}$$

where, according to (12) and (14),

$$\mathbf{S}_k = \begin{cases} \mathbf{B}_{(r_k)}^{(s_k)} (\mathbf{B}_{(r_k)}^{(s_k)\prime} \mathbf{W}_k \mathbf{B}_{(r_k)}^{(s_k)} + \lambda \boldsymbol{\Lambda})^{-1} \mathbf{B}_{(r_k)}^{(s_k)\prime} \mathbf{W}_k {\mathbf{D}_k}^{-1}, & s_k \leq v \\ \mathbf{B}^{*(s_k)} (\mathbf{B}^{*(s_k)\prime} \mathbf{W}_k \mathbf{B}^{*(s_k)} + \lambda_P \boldsymbol{\Delta}_2' \boldsymbol{\Delta}_2)^{-1} \mathbf{B}^{*(s_k)\prime} \mathbf{W}_k {\mathbf{D}_k}^{-1}, & s_k > v. \end{cases} \tag{18}$$

From (18), it becomes apparent that the type of the update of the chosen component depends on the presence or absence of a monotonicity restriction. Using the indicator function $I(.)$, (17) can be written in the closed form

$$\mathbf{B}^{(s)} \hat{\boldsymbol{\alpha}}_s^{(k)} = \mathbf{B}^{(s)} \hat{\boldsymbol{\alpha}}_s^{(k-1)} + I(s_k = s) \mathbf{S}_k (\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k-1)}).$$

With the approximation of the hat-matrix, one has $\hat{\boldsymbol{\mu}}^{(k-1)} \approx \mathbf{H}_{k-1}\mathbf{y}$, which leads to

$$\mathbf{B}^{(s)} \hat{\boldsymbol{\alpha}}_s^{(k)} \approx \mathbf{B}^{(s)} \hat{\boldsymbol{\alpha}}_s^{(k-1)} + I(s_k = s) \mathbf{S}_k (\mathbf{I} - \mathbf{H}_{k-1}) \mathbf{y},$$

and hence, in a recursive fashion,

$$\mathbf{B}^{(s)} \hat{\boldsymbol{\alpha}}_s^{(l)} \approx \mathbf{Q}_l^{(s)},$$

where

$$\mathbf{Q}_l = \sum_{k=1}^{l} I(s_k = s) \mathbf{S}_k (\mathbf{I} - \mathbf{H}_{k-1}).$$

Approximate confidence intervals for the estimate of the smooth component f_s after l boosting iterations are then found from

$$\widehat{cov}(\mathbf{Q}_l^{(s)} \mathbf{y}) = \mathbf{Q}_l^{(s)} \widehat{cov}(\mathbf{y}) \mathbf{Q}_l^{(s)\prime},$$

where $\widehat{cov}(\mathbf{y}) = diag(\hat{\sigma}_1^2, \dots, \hat{\sigma}_n^2)$.

4 Air pollution in São Paulo

In the following the air pollution data from Section 1 are investigated more closely. The objective is to evaluate the association between mortality of children under five attributed to respiratory causes and the concentration of SO_2, CO, PM_{10} and O_3. The response variable is the number of daily deaths attributed to respiratory causes in the city of São Paulo. The sample size is $n = 1067$. A standard approach for data of this type is to use a generalized additive 'core' model which includes terms to control for trend, seasonality and other influential variables like temperature or humidity, cf. Schwartz (1994). As the dependent variable consists of count data, we use a Poisson model along with the (natural) log-link, and consider the core model of Singer, Andre, Lima & Conçeicão (2002),

$$\begin{aligned}\eta = \log[E(\text{resp. deaths})] \quad &= \quad \alpha_0 + f_1(\text{time}) + f_2(\text{temp}) + f_3(\text{humidity}) \\ &\quad\; +\beta_1 \cdot \text{Monday} + \cdots + \beta_6 \cdot \text{Saturday} \\ &\quad\; +\beta_7 \cdot \text{non-resp. deaths}.\end{aligned} \tag{19}$$

The model includes non-specified functions to control for long-term seasonality (days), temperature (daily minimum, lag 2) and relative humidity (lag 0). In addition day of week dummies are included to control for short-term seasonality and the number of deaths by non-respiratory causes as a linear term. The basic strategy to investigate the effect of a specific pollutant is to take only this pollutant into the model. In the following, we will exemplarily focus on the concentration of SO_2, given in daily 24-h mean values of $\mu g/m^3$, considering the predictor

$$\tilde{\eta} = \eta + f_4(\text{SO}_2). \tag{20}$$

Since an increase in respiratory mortality with rising pollutant concentrations is expected, it is sensible to assume the function f_4 to be isotonic. To account for that assumption, model (20) has been fitted by the boosting procedure described in Section 2 (GMBBoost, Generalized Monotonic B-spline Boosting), where f_4 was estimated under the monotonicity constraint. We used a B-spline basis of degree 3 with $m = 20$ equidistant interior knots for each of the smooth components. The penalty parameter λ_P in the one step P-spline estimates from (14) for the non-monotonic components was set to 200. For the monotonic estimation of f_4, we chose a smaller λ of 20 due to the multiplication by weighting the penalty, see (6). It should be noted that the choice of penalty parameters is not crucial in boosting approaches; it is chosen for convenience such that the number of iterations is not too high. The fixed effects were re-estimated in each iteration. To stabilize the approximation of the hat-matrix, we put an additional shrinkage factor of $\nu = 0.1$ on the estimates in each boosting step. Boosting was stopped by AIC as well as by BIC. For comparison, we also fitted a generalized additive model using `gam()` from the R package `mgcv` (for details see R Foundation for Statistical Computing 2004 and Wood 2000).

Figure 2 shows the estimated curves $\hat{f}_1, \ldots, \hat{f}_4$ for the various fitting procedures. It is seen from $\hat{f}_1$ (upper left panel) that the seasonal pattern in mortality is evident for all three fitting methods. Mortality tends to decrease as temperature increases (upper right panel). Interestingly, the GAM fit yields a almost straight line, while both boosting estimates show a plateau between 10 and 15 °C, a result that has also been reported by Conceição, Miraglia, Kishi, Saldiva & Singer (2001). For the relative humidity (lower left panel), GAM results in an increasing, again almost straight line. AIC-stopped boosting shows two troughs at 65% and 90% relative humidity, while BIC-stopped boosting assigns only a marginal relevance to that component. The most interesting result is found in the fit for the concentration of SO_2 (lower right panel), where the monotonicity constraint is set in GMBBoost. The GAM fit shows the same effect that is seen in the simple introductory example presented in Figure 1: the curve is severely pulled down by the sparse points of high concentrations, resulting in the implausible result of decreasing mortality for concentrations larger than 40 $\mu g/m^3$. This phenomenon has been also detected by Einbeck, Andre & Singer (2004). Instead, GMBBoost shows a quite different behaviour. Since monotonicity is assumed for this component, one obtains a monotonic increasing fit, which remains constant on a high level of mortality for high pollutant concentrations. This result is in accordance with biological theory. For BIC-stopped boosting, the effect is flatter then for AIC-stopped boosting.

In Table 1, the parameter estimates for the fixed effects, controlling for long-term seasonality and non-respiratory deaths, are given for the different fitting methods, along with the corresponding values of AIC and BIC. It is seen that the estimates are rather stable across fitting procedures. More importantly, it is seen that AIC-stopped GMBBoost outperforms GAM distinctly in terms of the AIC criterion, indicating that the constrained boosting approach results in a more appropriate model for the present data. Interestingly, also BIC-stopped boosting does slightly better than GAM in terms of AIC. A similar result is seen for the BIC. The corresponding GMBBoost estimate performs best for this criterion, whereas GAM does even worse than AIC-stopped boosting. Since in the BIC criterion, the complexity of the fit is penalized stronger as compared to AIC, GMBBoost stops earlier for the former ($l_{opt}^{BIC} = 88$) than for the latter ($l_{opt}^{AIC} = 213$).

Figure 3 shows the curves fitted by AIC-stopped GMBBoost, and approximate 0.95 pointwise confidence bands as derived in Section 3. In the lights of confidence intervals, it is seen that the effects of temperature and humidity are rather weak.

In studies of the type presented here, one is often interested in the risk of death at a certain pollutant concentration, relative to the risk of death at the minimum concentration of that pollutant, see e.g. Singer, Andre, Lima & Conçeicão (2002) or Einbeck, Andre & Singer (2004). Let $SO_2(i)$ be the recorded concentration in observation i, $i = 1, \ldots, 1067$, and $SO_2(\min)$ the minimum concentration

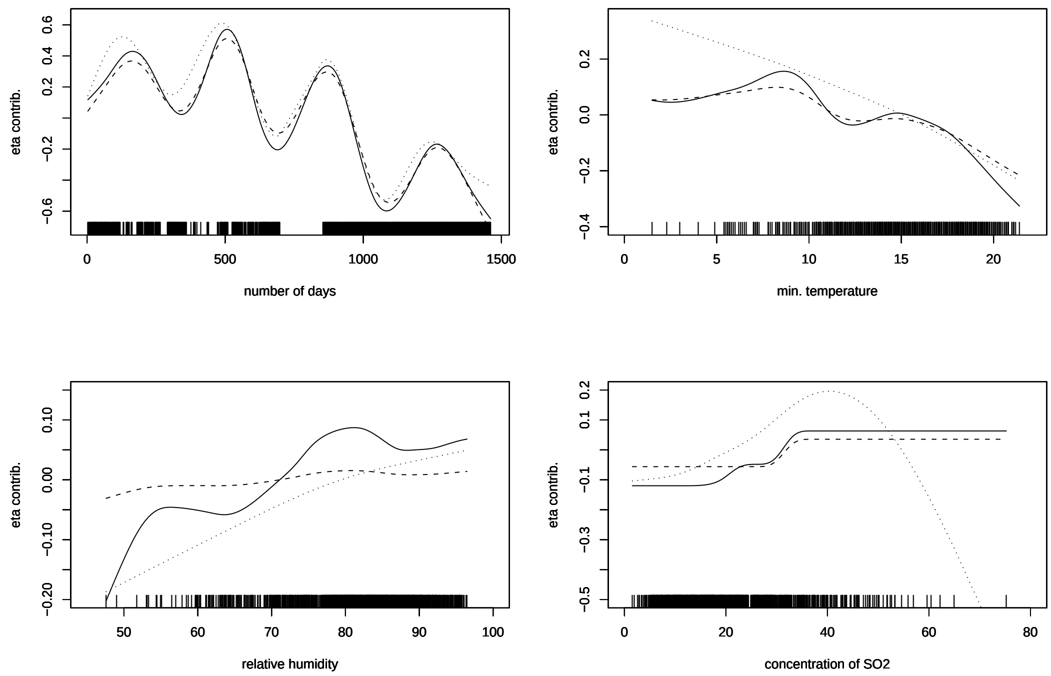


FIGURE 2: *Core model + $f_4(SO_2)$, estimated curves for the smooth components, GMBBoost with monotonic fitting of $f_4(SO_2)$, AIC-stopped (solid), BIC-stopped (dashed) and GAM (dotted). Data points are given as rug at the foot of each panel.*

recorded, then the relative risk of death is defined by

$$\begin{aligned} RR(i) &= \frac{E(\text{respiratory death}|\text{SO}_2(i))}{E(\text{respiratory death}|\text{SO}_2(\text{min}))} = \frac{\exp[\eta + f_4(\text{SO}_2(i))]}{\exp[\eta + f_4(\text{SO}_2(\text{min}))]} \\ &= \exp[f_4(\text{SO}_2(i)) - f_4(\text{SO}_2(\text{min}))]. \end{aligned}$$

In Figure 4, the estimated relative risk curve is given for the three fitting methods. The implausible result of the GAM fit which indicates that high SO_2-concentration causes a decrease in the relative risk of death, is apparent. In contrast, the GMBBoost fits show a monotonic increase of the risk curve for values up to 35 $\mu g/m^3$. For larger concentrations, the risk remains at a fairly high level for the AIC-stopped boosting, whereas the effect is not as strong if boosting is stopped by BIC.

5 Concluding remarks

A procedure is proposed that allows to use the information on monotonicity for one or more components within a generalized additive model. By using

	GAM	GMBBoost, AIC	GMBBoost, BIC
intercept	0.9687	1.0971	1.0995
Monday	-0.1578	-0.1621	-0.1799
Tuesday	-0.2094	-0.2347	-0.2717
Wednesday	-0.0034	-0.0048	-0.0061
Thursday	-0.1213	-0.1136	-0.1105
Friday	-0.1673	-0.1733	-0.1727
Saturday	-0.1123	-0.1114	-0.1123
non-resp. deaths	-0.0088	-0.0086	-0.0058
AIC	1292.0622	1279.6322	1289.8215
BIC	1408.3948	1393.3433	1375.8885

TABLE 1: *Core model + $f_4(SO_2)$, estimates of fixed coefficients for GAM and GMBBoost.*

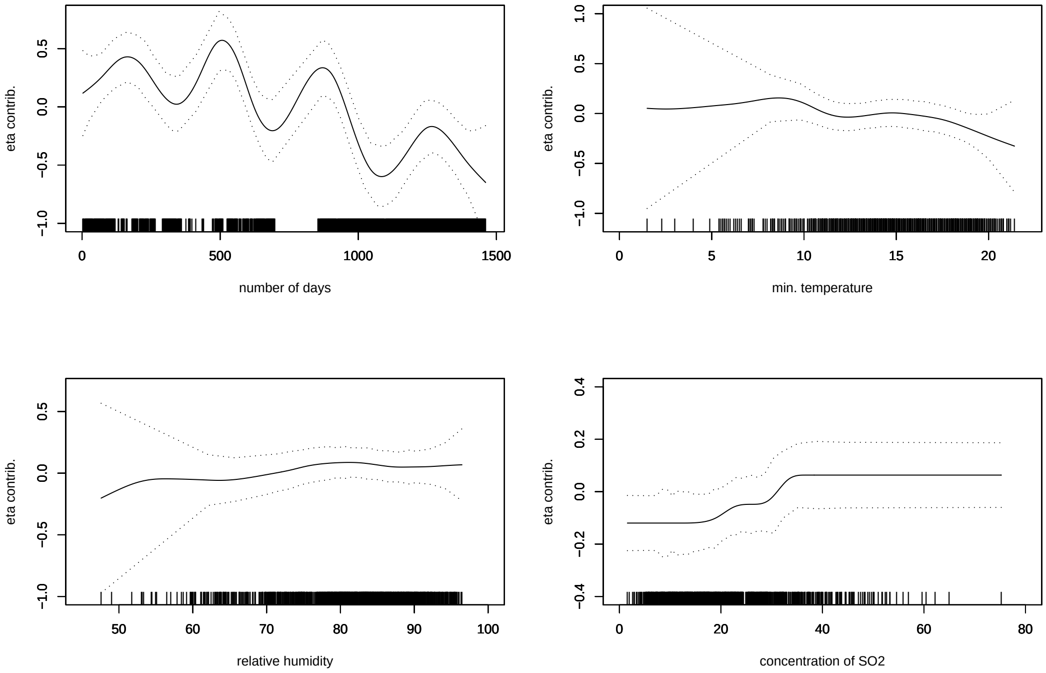


FIGURE 3: *Core model + $f_4(SO_2)$, AIC-stopped GMBBoost with monotonic fitting of $f_4(SO_2)$ along with the approximate confidence intervals. Data points are given as rug at the foot of each panel.*

monotonicity, the procedure prevents that few outlying observations yield implausible fits. These effects may be avoided to a certain degree by downweighting

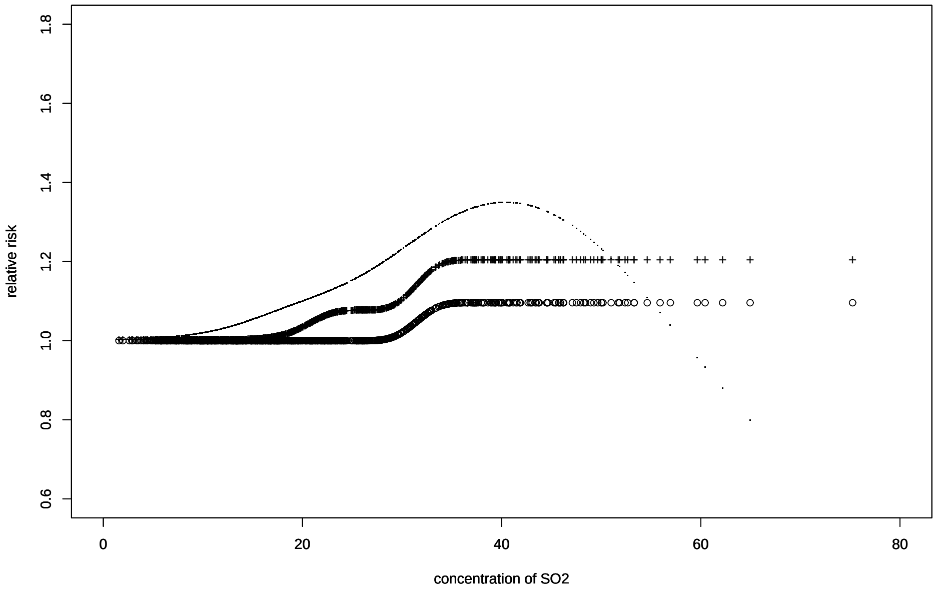


FIGURE 4: *Relative risk curves vs. SO_2 concentration for GAM ($\cdot$), GMBBoost, AIC-stopped (+) and BIC-stopped ($\circ$).*

observations with small design density (Einbeck, Andre & Singer 2004). However, with downweighting approaches, problems arise in higher dimensions, since densities have to be estimated. The monotone regression boosting approach does not suffer from these problems. It should also be noted that the problem of choosing smoothing parameters - which in case of higher dimensional covariates is hard to tackle - is avoided in boosting techniques. The only crucial tuning parameter is the number of boosting iterations, which is chosen by the AIC or BIC criterion.

Acknowledgements

We gratefully acknowledge support from the Deutsche Forschungsgemeinschaft (SFB 386, "Statistical Analysis of Discrete Structures"). We thank Jochen Einbeck and Julio M. Singer for letting us use the data on air pollution in São Paulo.

References

BÜHLMANN, P. (2004). Boosting for high–dimensional linear models. Technical Report, ETH Zürich.

BÜHLMANN, P. AND YU, B. (2003). Boosting with the L_2loss: regression and classification. *Journal of the American Statistical Association 98*, 324–339.

BÜHLMANN, P. AND YU, B. (2005). Boosting, model selection, lasso and nonnegative garrote. Technical Report, ETH Zürich.

BREZGER, A. AND STEINER, W. J. (2004). Monotonic regression based on Bayesian P-splines: an application to estimating price response functions from store-level scanner data. SFB Discussion Paper 331, LMU München.

CONCEIÇÃO, G. M. S., MIRAGLIA, S. G. E. K., KISHI, H. S., SALDIVA, P. H. N., AND SINGER, J. M. (2001). Air pollution and children mortality: a time series study in São Paulo, Brazil. *Environmental Health Perspectives 109*, 347–350.

DE BOOR, C. (1978). *A Practical Guide to Splines.* New York: Springer-Verlag.

EILERS, P. H. C. AND MARX, B. D. (1996). Flexible smoothing with B-splines and Penalties. *Statistical Science 11*, 89–121.

EINBECK, J., ANDRE, C. D. S., AND SINGER, J. M. (2004). Local smoothing with robustness against outlying predictors. *Environmetrics 15*, 541–554.

FAN, J. AND GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications.* London: Chapman & Hall.

FRIEDMAN, J. H. AND TIBSHIRANI, R. (1984). The monotone smoothing of scatterplots. *Technometrics 26*, 243–250.

GREEN, D. J. AND SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* London: Chapman & Hall.

HASTIE, T. AND TIBSHIRANI, R. (1990). *Generalized Additive Models.* London: Chapman & Hall.

KELLY, C. AND RICE, J. (1991). Monotone smoothing with application to dose–response curves and the assessment of synergism. *Biometrics 46*, 1071–1085.

MAMMEN, E., MARRON, J., TURLACH, B., AND WAND, M. (2001). A general projection framework for constrained smoothing. *Statistical Science 16*(3), 232–248.

MARX, B., EILERS, P., AND SMITH, E. (1992). Ridge likelihood estimation for generalized linear regression. In R. van der Heijden, W. Jansen, B. Francis, & G. Seeber (Eds.), *Statistical Modelling*, pp. 227–238. Amsterdam: North-Holland.

MARX, D. B. AND EILERS, P. H. C. (1998). Direct generalized additive modelling with penalized likelihood. *Comp. Stat. & Data Analysis 28*, 193–209.

MCCULLAGH, P. AND NELDER, J. A. (1989). *Generalized Linear Models* (2nd ed.). New York: Chapman & Hall.

MUKERJEE, H. (1988). Monotone nonparametric regression. *The Annals of Statistics 16*, 741–750.

R Foundation for Statistical Computing (2004). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

RAMSAY, J. O. (1988). Monotone regression splines in action. *Statistical Science 3*, 425–461.

ROBERTSON, T., WRIGHT, F. T., AND DYKSTRA, R. L. (1988). *Order–Restricted Statistical Inference.* New York: Wiley.

SCHAPIRE, R. E. (1990). The strength of weak learnability. *Machine Learning 5*, 197–227.

SCHWARTZ, J. (1994). Nonparametric smoothing in the analysis of air pollution and respiratory illness. *Canadian Journal of Statistics 22*(4), 471–487.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*, 461–464.

SINGER, J. M., ANDRE, C. D. S., LIMA, P. L., AND CONÇEICÃO, G. M. S. (2002). Associaton between atmospheric pollution and mortality in São Paulo, Brazil: regression models and analysis strategy. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L1 Norm and Related Methods*, pp. 439–450. Berlin: Birkhäuser.

SPECKMAN, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society B 50*, 413–436.

THEIL, H. AND GOLDBERGER, A. S. (1961). On pure and mixed estimation in econometrics. *International Economic Review 2*, 65–78.

TUTZ, G. AND BINDER, H. (2004). Generalized additive modelling with implicit variable selection by likelihood based boosting. SFB Discussion Paper 401, LMU München.

TUTZ, G. AND LEITENSTORFER, F. (2005). Generalized smooth monotonic regression. SFB Discussion Paper 417, LMU München.

WOOD, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society B 62*, 413–428.