*Biostatistics* (2007), **8**, 3, *pp.* 654–673
doi:10.1093/biostatistics/kxl036
Advance Access publication on October 24, 2006

# Generalized monotonic regression based on B-splines with an application to air pollution data

FLORIAN LEITENSTORFER*, GERHARD TUTZ

*Department of Statistics, Ludwig-Maximilians-Universität München,*
*Akademiestraße 1, 80799 München, Germany*
florian.leitenstorfer@stat.uni-muenchen.de

## Summary

In many studies, it is known that one or more of the covariates have a monotonic effect on the response variable. In these circumstances, standard fitting methods for generalized additive models (GAMs) generate implausible results. A fitting procedure is proposed that incorporates monotonicity assumptions on one or more smooth components within a GAM framework. The 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 the 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 is investigated.

*Keywords*: Air pollution data; Generalized additive models; Likelihood-based boosting; Monotonic regression.

## 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 has 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, 1994b, or Conceição *and others*, 2001). In these analyses, an increase of deaths or cases of illness is expected with an increasing concentration of a specific pollutant. When standard smoothing techniques, like spline smoothing (Green and Silverman, 1994) or local polynomial fitting (Fan and Gijbels, 1996), are applied to data of this type in a generalized additive modeling approach, the fitted curves may lead to unconvincing results. In the following, it is proposed to incorporate knowledge about monotonic relationships in the estimation by using monotonic regression methods. Other important biometrical problems that require monotonic regression techniques are the estimation of dose–response functions (e.g. Kelly and Rice, 1991; Lee, 1996, or Dilleen *and others*, 2003) and the estimation of human growth curves (e.g. Ducharme and Fontez, 2004).

Starting from the pool adjacent violators algorithm (PAVA) (see, e.g. Robertson *and others*, 1988) which produces a step function, a variety of methods have been developed to smooth the PAVA results, so as to obtain a smooth estimate of the underlying monotonic function. Details of such approaches, which

*To whom correspondence should be addressed.

are mainly based on kernel regression techniques, are given in Friedman and Tibshirani (1984), Mukerjee (1988), or Mammen *and others* (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, that is, $f = \sum_j \alpha_j B_j$. To assure monotonicity of the estimate, constraints have to be put on the coefficients $\alpha_j$. Ramsay (1988) suggests the use of monotonic basis functions (integrated splines), while Kelly and Rice (1991) propose a B-spline basis. As the B-spline approach has become very popular in nonparametric regression (see Eilers and Marx, 1996), we will focus on the latter.

Many publications on monotonic regression focus on unidimensional smoothing problems with a Gaussian response variable $y$ (see, e.g. Ramsay, 1998; Zhang, 2004, or Turlach, 2005). 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 and Rice, 1991), which lead to unsatisfactory 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 and Yu, 2003). As demonstrated by Tutz and Leitenstorfer (2006), 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 and Leitenstorfer (2006) 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 are computationally burdensome and often yield unstable estimates. From a Bayesian perspective, a B-spline approach to monotonic regression has been suggested by Brezger and Steiner (2004).

We illustrate generalized monotonic regression techniques on a data set that has previously been analyzed by Conceição *and others* (2001), Singer *and others* (2002), and Einbeck *and others* (2004). The data have been collected to evaluate the association between mortality due to respiratory causes and the concentration of various air pollutants in the city of São Paulo, Brazil, from 1994 to 1997. The analysis presented here focuses on the effect on people older than 65 years, which is one of the most susceptible population segments. GAMs are frequently used for the analysis of such data, where one first aims at controlling for seasonal and weather effects and then includes an air pollution variable in the model (see Schwartz, 1994b). There are numerous publications that suggest that respiratory mortality increases with the concentration of air pollutants (see, e.g. Schwartz, 1994a, for a review). Under this assumption, it seems sensible to estimate smooth effects of pollution concentration on mortality under a monotonicity constraint, resulting in more reliable fits.

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, the performance of our approach is evaluated in various simulation studies. In Section 4, we take a closer look on the data set mentioned above. Note that throughout the paper, we take monotonic to mean nondecreasing.

## 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 GLMs (e.g. McCullagh and 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 $\theta_i$ is the canonical parameter and $\phi$ 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 GLMs, $\eta(x)$ 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

$$f(x) \geqslant f(z) \quad \text{if } x > z. \tag{2.1}$$

Obviously, monotonicity in $\eta$ 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 and 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

$$\eta(x) = \alpha_0 + \sum_{j=1}^{m} \alpha_j B_j(x, q), \tag{2.2}$$

where $q$ denotes the degree of the B-splines and $m = \tilde{m} - 1 + q$ (the number of basis functions). 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 $q \geqslant 1$. 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 $B_j(x, q-1) \geqslant 0$, it follows from

$$\alpha_{j+1} \geqslant \alpha_j \tag{2.3}$$

that $\eta'(x) \geqslant 0$ holds. In other words, since (2.3) is a sufficient condition for the monotonicity of $\eta(x)$, the sequence of coefficients $\alpha_j$ has to be nondecreasing in order to obtain monotonic functions. This property of B-splines has been previously exploited by Kelly and Rice (1991) and Brezger and Steiner (2004) in a monotonic regression setting. Note that throughout this paper, we work with B-splines that are centered by their corresponding integral.

### 2.2 *An outline of the algorithm*

Boosting was originally 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 and Yu, 2003; Bühlmann, 2006). 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 and 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 $\mathbf{y} = (y_1, \ldots, y_n)'$ and $\mathbf{x} = (x_1, \ldots, x_n)'$. Based on the expansion into basis functions, 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 with predictor

$$\eta(x_i) = \alpha_{0(r)} + \alpha_{1(r)}\left(\sum_{j=1}^{r} B_j(x_i)\right) + \alpha_{2(r)}\left(\sum_{j=r+1}^{m} B_j(x_i)\right). \tag{2.4}$$

When fitting model (2.4), the monotonicity constraint is easily checked by comparing the estimates $\hat{\alpha}_{1(r)}$ and $\hat{\alpha}_{2(r)}$, since monotonicity follows from $\hat{\alpha}_{2(r)} \geqslant \hat{\alpha}_{1(r)}$. Given an estimate from the previous fitting

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

refitting is performed by

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

It is obvious that $\hat{\eta}_{\text{new}}$ is monotonic if the estimates fulfill $\hat{\alpha}_{2(r)} \geqslant \hat{\alpha}_{1(r)}$, provided that the previous estimate $\hat{\eta}_{\text{old}}$ was monotonic. The grouping of basis functions into $B_1, \ldots, B_r$ and $B_{r+1}, \ldots, B_m$, the effect of which is adapted by the amount $\alpha_{1(r)}$ in the first and $\alpha_{2(r)}$ in the second group, allows monotonicity to be controlled in a simple way. Fitting a full model with $m$ new parameters would imply much more computational effort. Furthermore, if individual constraints are imposed on each of the $m$ parameters, problems with convergence might occur. Instead, the possible groupings ($r = 1, \ldots, m-1$) are evaluated and by analogy with 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 in linear models (see Theil and Goldberger, 1961). In the usual form of a smoothed estimate based on B-splines, model (2.4) is given by

$$\eta(x_i) = \alpha_{0(r)} + \sum_{j=1}^{r} \alpha_j B_j(x_i) + \sum_{j=r+1}^{m} \alpha_j B_j(x_i) \tag{2.5}$$

with the constraints $\alpha_1 = \cdots = \alpha_r = \alpha_{1(r)}$, $\alpha_{r+1} = \cdots = \alpha_m = \alpha_{2(r)}$. 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 and Binder (2006), the fit of model (2.4) is embedded into the framework of penalized likelihood estimation. Moreover, the model is generalized to a model that contains parametric effects in addition to the smooth monotonic effects. Thus, the intercept term is replaced by $\mathbf{z}'\boldsymbol{\alpha}_0$, where $\mathbf{z}$ is a vector of covariates and $\boldsymbol{\alpha}_\mathbf{0}$ is an unknown parameter vector (possibly specifying only the intercept). It is assumed that $\mathbf{z}_i$ always contains an intercept. In order to avoid identifiability issues which arise for B-splines in connection with an intercept, the update step is split up into 2 parts. In the first part, the smooth component is updated and in the second, the parametric term. In the first part, one fits by penalized likelihood. Therefore, one considers

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

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

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

where $l_i(\boldsymbol{\alpha}_{(r)}) = l_i(h(\mathbf{B}_{(r)}\boldsymbol{\alpha}_{(r)}))$ is the usual log-likelihood contribution of the $i$th observation and $P(\boldsymbol{\alpha}_{(r)}) = (\lambda/2)\boldsymbol{\alpha}'_{(r)}\boldsymbol{\alpha}_{(r)}$ represents the penalty term with ridge parameter $\lambda$. Model (2.5) is asymmetric in a specific sense. If, for example, $r = 2$, the first constraint $\alpha_1 = \alpha_2$ concerns only 2 parameters, whereas the second constraint $\alpha_3 = \cdots = \alpha_m$ concerns $m - 2$ parameters, which for $m = 22$ means 20 parameters are restricted. To account for this asymmetry, we tried the modified penalty $P(\boldsymbol{\alpha}_{(r)}) = \frac{\lambda}{2m}(r\alpha_{1(r)}^2 + (m - r)\alpha_{2(r)}^2)$, where the parameters are weighted by the number of parameters that are implicitly considered as identical. However, numerical examples showed that the performance of the modified penalty is nearly the same as for the usual ridge penalty $P(\boldsymbol{\alpha}_{(r)}) = \frac{\lambda}{2}(\alpha_{1(r)}^2 + \alpha_{2(r)}^2)$. Hence, in the following we use this simpler scheme, which in matrix form leads to the penalized log-likelihood

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

where $\boldsymbol{\Lambda} = \mathrm{diag}\{1, 1\}$ and $\lambda > 0$ represents the ridge parameter. Derivation yields the corresponding penalized score function

$$s_{\mathrm{p}}(\boldsymbol{\alpha}_{(r)}) = \frac{\partial l_{\mathrm{p}}(\boldsymbol{\alpha}_{(r)})}{\partial \boldsymbol{\alpha}_{(r)}} = \mathbf{B}'_{(r)}\mathbf{W}(\boldsymbol{\eta})\mathbf{D}(\boldsymbol{\eta})^{-1}(\mathbf{y} - h(\boldsymbol{\eta})) - \lambda\boldsymbol{\Lambda}\boldsymbol{\alpha}_{(r)}, \tag{2.6}$$

with $\mathbf{W}(\boldsymbol{\eta}) = \mathbf{D}^2(\boldsymbol{\eta})\boldsymbol{\Sigma}(\boldsymbol{\eta})^{-1}$, $\mathbf{D}(\boldsymbol{\eta}) = \mathrm{diag}\{\partial h(\eta_1)/\partial\eta, \ldots, \partial h(\eta_n)/\partial\eta\}$, $\boldsymbol{\Sigma}(\boldsymbol{\eta}) = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$, $\sigma_i^2 = \mathrm{var}(y_i)$, all of them evaluated at the current value of $\eta$. The monotonicity constraint from (2.3) is incorporated by taking into account only estimates which fulfill $\hat{\alpha}_{2(r)} \geqslant \hat{\alpha}_{1(r)}$. It is easily seen that the update scheme given below yields the desired nondecreasing sequences of estimates $\hat{\alpha}_1, \ldots, \hat{\alpha}_m$ in each boosting iteration. The update of the parametric term $\mathbf{z}'\boldsymbol{\alpha}_0$ is performed in the same way but without penalization and with the design matrix determined by $\mathbf{Z} = (\mathbf{1}, \mathbf{z}_1, \ldots, \mathbf{z}_u)$, where $\mathbf{z}_1, \ldots, \mathbf{z}_u$ denote the observed covariate values.

---

**Monotonic Likelihood Boosting for B-splines**

**Step 1:** (Initialization)

Set $\hat{\boldsymbol{\alpha}}_0^{(0)} = (g(\bar{y}), 0, \ldots, 0)'$, $\hat{\boldsymbol{\alpha}}^{(0)} = (0, \ldots, 0)'$, $\hat{\boldsymbol{\eta}}^{(0)} = (g(\bar{y}), \ldots, g(\bar{y}))'$, and $\hat{\boldsymbol{\mu}}^{(0)} = (\bar{y}, \ldots, \bar{y})'$.

**Step 2:** (Iteration)

For $l = 1, 2, \ldots$

1) *Fitting step, monotone component*

For $r = 1, \ldots, m - 1$, compute the modified ridge estimate based on the 1-step Fisher scoring,

$$\hat{\boldsymbol{\alpha}}_{(r)} = (\mathbf{B}'_{(r)}\mathbf{W}_l\mathbf{B}_{(r)} + \lambda\boldsymbol{\Lambda})^{-1}\mathbf{B}'_{(r)}\mathbf{W}_l\mathbf{D}_l{}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(l-1)}), \tag{2.7}$$

where $\hat{\boldsymbol{\alpha}}_{(r)} = (\hat{\alpha}_{1(r)}, \hat{\alpha}_{2(r)})'$, $\mathbf{W}_l = \mathbf{W}(\hat{\boldsymbol{\eta}}^{(l-1)})$, $\mathbf{D}_l = \mathbf{D}(\hat{\boldsymbol{\eta}}^{(l-1)})$, and $\hat{\boldsymbol{\mu}}^{(l-1)} = h(\hat{\boldsymbol{\eta}}^{(l-1)})$. Let $A = \{r\colon \hat{\alpha}_{1(r)} \leqslant \hat{\alpha}_{2(r)}\}$ denote the candidates that fulfill the monotonicity constraint. If $A = \emptyset$, stop. Otherwise continue with Step 2.

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

$$r_l = \arg\min_{r \in A} \text{Dev}(\tilde{\boldsymbol{\eta}}_{(r),\text{new}}),$$

where $\text{Dev}(\tilde{\boldsymbol{\eta}}_{(r),\text{new}}) = 2\sum_{i=1}^{n}[l_i(y_i) - l_i(h(\tilde{\eta}_{i,(r),\text{new}}))]$. Set

$$\hat{\alpha}_j^{(l)} = \begin{cases} \hat{\alpha}_j^{(l-1)} + \hat{\alpha}_{1(r_l)}, & 1 \leqslant j \leqslant r_l, \\ \hat{\alpha}_j^{(l-1)} + \hat{\alpha}_{2(r_l)}, & j > r_l, \end{cases} \tag{2.8}$$

$$\tilde{\boldsymbol{\eta}}^{(l-1)} = \hat{\boldsymbol{\eta}}^{(l-1)} + \mathbf{B}_{(r_l)}\hat{\boldsymbol{\alpha}}_{(r_l)}, \quad \text{and} \quad \tilde{\boldsymbol{\mu}}^{(l-1)} = h(\tilde{\boldsymbol{\eta}}^{(l-1)}).$$

3) *Fitting step and update, parametric term*

Based on the 1-step Fisher scoring, one obtains

$$\hat{\boldsymbol{\alpha}}_0 = (\mathbf{Z}'\tilde{\mathbf{W}}_l\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{W}}_l\tilde{\mathbf{D}}_l^{-1}(\mathbf{y} - \tilde{\boldsymbol{\mu}}^{(l-1)}),$$

where $\tilde{\mathbf{W}}_l = \mathbf{W}(\tilde{\boldsymbol{\eta}}^{(l-1)})$ and $\tilde{\mathbf{D}}_l = \mathbf{D}(\tilde{\boldsymbol{\eta}}^{(l-1)})$. Set

$$\hat{\boldsymbol{\alpha}}_0^{(l)} = \hat{\boldsymbol{\alpha}}_0^{(l-1)} + \hat{\boldsymbol{\alpha}}_0, \quad \hat{\boldsymbol{\eta}}_0^{(l)} = \tilde{\boldsymbol{\eta}}_0^{(l-1)} + \mathbf{Z}\hat{\boldsymbol{\alpha}}_0^{(l)}, \quad \text{and} \quad \hat{\boldsymbol{\mu}}^{(l)} = h(\hat{\boldsymbol{\eta}}^{(l)}).$$

---

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 and Yu (2003) and Bühlmann (2006) 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 and Leitenstorfer (2006) 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 and Leitenstorfer (2006). With $\mathbf{M}_0 = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$, $\mathbf{M}_l = \boldsymbol{\Sigma}_l^{1/2}\mathbf{W}_l^{1/2}\mathbf{B}_{(r_l)}\big(\mathbf{B}_{(r_l)}'\mathbf{W}_l\mathbf{B}_{(r_l)} + \lambda\boldsymbol{\Lambda}\big)^{-1}\mathbf{B}_{(r_l)}'\mathbf{W}_l^{1/2}\boldsymbol{\Sigma}_l^{-1/2}$, where $\mathbf{W}_l = \mathbf{W}(\hat{\boldsymbol{\eta}}^{(l-1)})$ and $\boldsymbol{\Sigma}_l = \boldsymbol{\Sigma}(\hat{\boldsymbol{\eta}}^{(l-1)})$, and $\tilde{\mathbf{M}}_l = \tilde{\boldsymbol{\Sigma}}_l^{1/2}\tilde{\mathbf{W}}_l^{1/2}\mathbf{Z}(\mathbf{Z}'\tilde{\mathbf{W}}_l\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{W}}_l^{1/2}\tilde{\boldsymbol{\Sigma}}_l^{-1/2}$, where $\tilde{\mathbf{W}}_l = \mathbf{W}(\tilde{\boldsymbol{\eta}}^{(l-1)})$ and $\tilde{\boldsymbol{\Sigma}}_l = \boldsymbol{\Sigma}(\tilde{\boldsymbol{\eta}}^{(l-1)})$, $l = 1, 2, \dots$, the approximate hat matrix is given by

$$\mathbf{H}_l = \mathbf{I} - \left[\prod_{j=1}^{l}(\mathbf{I} - \tilde{\mathbf{M}}_{l-j+1})(\mathbf{I} - \mathbf{M}_{l-j+1})\right](\mathbf{I} - \mathbf{M}_0), \tag{2.9}$$

with $\hat{\boldsymbol{\mu}}^{(l)} \approx \mathbf{H}_l\mathbf{y}$. By considering $\text{tr}(\mathbf{H}_l)$ as the degree of freedom of the smoother, we use as potential stopping criteria an information criterion proposed by Akaike (AIC) and the bayesian information criterion (BIC) criteria, $\text{AIC}(l) = \text{Dev}_l + 2\text{tr}(\mathbf{H}_l)$ and $\text{BIC}(l) = \text{Dev}_l + \log(n)\text{tr}(\mathbf{H}_l)$, where $\text{Dev}_l = 2\sum_{i=1}^{n}[l_i(y_i) - l_i(h(\hat{\eta}_i^{(l)}))]$ denotes the deviance of the model in the $l$th boosting step. The optimal number of boosting iterations is defined by $l_{\text{opt}}^{\text{AIC}} = \arg\min_l \text{AIC}(l)$ or $l_{\text{opt}}^{\text{BIC}} = \arg\min_l \text{BIC}(l)$. Since the BIC (Schwarz, 1978) penalizes the complexity of the fit stronger, usually sparser models result. A more extensive treatment of stopping criteria for boosting algorithms is given in Bühlmann and Yu (2006).

### 2.3 *Extension to GAMs*

In biometrical or ecological problems, one is usually interested in the effect of several smooth predictor variables, some of which might have monotonic influence on $y$. In the following, we demonstrate that the concept given above can easily be extended to a GAM setting (see, e.g. Hastie and Tibshirani, 1990, or Marx and Eilers, 1998). Let

$$\eta = \mathbf{z}'\boldsymbol{\alpha}_0 + \sum_{s=1}^{p} f_s(x_s), \tag{2.10}$$

where for some of the $p$ unknown smooth functions (say $f_1, \ldots, f_v, v \leqslant p$) monotonicity constraints are assumed to hold. Using the matrix notation from above, we have a design matrix $(\mathbf{Z}, \mathbf{X})$, with the matrix of linear terms $\mathbf{Z}$ and the matrix of smooth components $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$, where $\mathbf{x}_s = (x_{1s}, \ldots, x_{ns})'$. Componentwise expansion of $\mathbf{X}$ into B-spline basis functions leads to the matrix $(\mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(p)})$, where $\mathbf{B}^{(s)}$ refers to the $s$th predictor.

It is essential to distinguish between components that are or are not under monotonicity restrictions. For the former, grouping of basis functions is done within each component in the same way as described in (2.4). For the unconstrained components, we follow Bühlmann and Yu (2003) and Tutz and Binder (2006) and use penalized regression splines (P-splines, cf. Eilers and Marx, 1996) as weak learners 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 all smooth terms in the model is then given by $\boldsymbol{\alpha} = (\alpha_{11}, \ldots, \alpha_{1m}, \ldots, \alpha_{p1}, \ldots, \alpha_{pm})'$. Thus, Step 2 (iteration) of the algorithm described above is extended as follows:

**Step 2:** (Iteration)

For $l = 1, 2, \ldots$

1) *Fitting step, smooth components*
For $s = 1, \ldots, p$,

- If $s \in \{1, \ldots, v\}$ (the components under monotonicity constraint), compute the estimates from (2.7) componentwise for the possible groupings $r = 1, \ldots, m-1$, with

$$\mathbf{B}^{(s)}_{(r)} = \mathbf{B}^{(s)}\mathbf{R}_{(r)}. \tag{2.11}$$

The set of indices for components $s$ and split points $r$ that satisfy the monotonicity constraint is given by

$$A_1 = \{(s, r) \in \{1, \ldots, v\} \times \{1, \ldots, (m-1)\}\colon \hat{\alpha}^{(s)}_{1(r)} \leqslant \hat{\alpha}^{(s)}_{2(r)}\}.$$

- If $s \in \{v+1, \ldots, p\}$ (the components without constraints), compute the 1-step Fisher scoring estimate of the P-spline,

$$\hat{\boldsymbol{\alpha}}^{(s)} = (\mathbf{B}^{(s)\prime}\mathbf{W}_l\mathbf{B}^{(s)} + \lambda_{\mathrm{P}}\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2)^{-1}\mathbf{B}^{(s)\prime}\mathbf{W}_l\mathbf{D}_l{}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(l-1)}), \tag{2.12}$$

where

$$\boldsymbol{\Delta}_2 = \begin{pmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{pmatrix}$$

denotes the matrix representation of the second-order differences. Since the P-spline fit (2.12) 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 $A_2 = \{(s, 0), s \in \{\upsilon + 1, \ldots, p\}\}$, yielding

$$A = A_1 \cup A_2. \tag{2.13}$$

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

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

In each iteration, only the basis coefficients belonging to the chosen component $s_l$ are refitted. That is, if the selected $s_l$ is in $\{1, \ldots, \upsilon\}$, then $\hat{\alpha}^{(l)}_{s_l,j}$, $j = 1, \ldots, m$, are updated by the refitting scheme (2.8). If $s_l > \upsilon$, then update $\hat{\alpha}^{(l)}_{s_l,j} = \hat{\alpha}^{(l-1)}_{s_l,j} + \hat{\alpha}^{(s_l)}_j$, with $\hat{\boldsymbol{\alpha}}^{(s_l)}$ from (2.12).

3) *Fitting step and update, parametric terms. See above.*

By using $\mathbf{B}^{(s)}_{(r)}$ from (2.11), along with $\mathbf{B}^{(s)}$ and the penalty matrices for the P-spline estimates, the hat-matrix approximation and the corresponding AIC and BIC stopping criteria can be extended to the additive setting. One obtains the hat matrix from (2.9), where

$$\mathbf{M}_l = \begin{cases} \boldsymbol{\Sigma}_l^{1/2}\mathbf{W}_l^{1/2}\mathbf{B}^{(s_l)}_{(r_l)}\big(\mathbf{B}^{(s_l)\prime}_{(r_l)}\mathbf{W}_l\mathbf{B}^{(s_l)}_{(r_l)} + \lambda\boldsymbol{\Lambda}\big)^{-1}\mathbf{B}^{(s_l)\prime}_{(r_l)}\mathbf{W}_l^{1/2}\boldsymbol{\Sigma}_l^{-1/2}, & s_l \leqslant \upsilon, \\[2ex] \boldsymbol{\Sigma}_l^{1/2}\mathbf{W}_l^{1/2}\mathbf{B}^{(s_l)}(\mathbf{B}^{(s_l)\prime}\mathbf{W}_l\mathbf{B}^{(s_l)} + \lambda_{\text{P}}\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2)^{-1}\mathbf{B}^{(s_l)\prime}\mathbf{W}_l^{1/2}\boldsymbol{\Sigma}_l^{-1/2}, & s_l > \upsilon, \end{cases}$$

with $l = 1, 2, \ldots$. 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 and 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.

A derivation of standard deviations for function and parameter estimates is given in Appendix A.1.

## 3. Simulation results

In order to evaluate the performance of the proposed method, we conduct some simulation studies. In a first setting, a unidimensional Poisson regression model is considered, with response $y_i$ generated from $\text{P}(\exp(\eta_i))$, where $\eta_i = \eta(x_i)$ is specified by a monotonic function. The $x_i$ are drawn from a $U[0, 5]$-distribution. We investigate a step function, $\eta(x) = 2cI(x > 2.5)$, and a plateau function, $\eta(x) = c(2/\{1 + \exp[-10(x-1)]\} + 2/\{1 + \exp[-5(x-4)]\} - 1)$. The strength of the signal is controlled by the constant $c$.

For Generalized Monotonic B-spline Boosting (GMBBoost), a B-spline basis of degree $q = 3$ is used, with 20 equally spaced knots in the domain of the data, which results in $m = 22$ basis functions. The ridge parameter has been chosen as $\lambda = 300$, and the maximum number of boosting iterations is limited to $L = 500$. Note that for better comparability, the predictor variable is always rescaled to [0, 1] before proceeding further.

GMBBoost is compared to unconstrained penalized regression splines as implemented in the R library "mgcv", where the penalization parameter is determined by the unbiased risk estimation (UBRE) criterion (see Wood, 2000, 2001, 2003). In order to obtain comparability with the proposed approach, a cubic regression spline basis with a dimension of $k = 22$ is used. Furthermore, following a suggestion of a referee, we consider a monotonicity-constrained version of this approach, based on quadratic programming. This involves embedding a monotone smoother in an iteratively reweighted least-squares loop. When using regression splines, this monotone smoother can be obtained by the function “pcls” in mgcv which solves a penalized least-squares quadratic programming problem. For details, see Wood (1994). The smoothing parameter is chosen by UBRE applied to the unconstrained model. Ordinary PAVA is also included, since in 1-parameter exponential families, it is the restricted maximum likelihood estimate for isotonic regression (see Robertson *and others*, 1988, Theorem 1.5.2).

A criterion for the performance of the fitting methods is the averaged Kullback–Leibler (AKL) distance,

$$\mathrm{AKL} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{KL}[\hat{\mu}_i, \mu_i], \tag{3.1}$$

where in the case of Poisson regression we have $\mathrm{KL}[\hat{\mu}_i, \mu_i] = \hat{\mu}_i \log(\frac{\hat{\mu}_i}{\mu_i}) - (\hat{\mu}_i - \mu_i)$, with $\hat{\mu}_i = \exp[\hat{\eta}(x_i)]$ and $\mu_i = \exp[\eta(x_i)]$. The means of AKL over $S = 250$ simulated data sets are given in Table 1 for selected sample sizes and noise levels. For the step function example, it is seen that GMBBoost is a strong competitor that clearly outperforms the unconstrained and constrained MGCV fits. Note that especially in the lower noise case, in about 10% of the simulated data sets, no fit could be obtained by the restricted version of MGCV. PAVA, which might be thought of being appropriate for this example due to its noncontinuous character, does better only in the case of a stronger signal and $n = 100$. For the plateau

Table 1. *AKL error over* $S = 250$ *simulated data sets for* 1*-dimensional Poisson regression, with corresponding standard errors. The number of instances where no fit could be obtained is given in brackets, the best* 2 *procedures in boldface*

| | | MGCV | monMGCV | PAVA | GMB (AIC) | GMB (BIC) |
|---|---|---|---|---|---|---|
| Step function | | | | | | |
| $c = 0.5$ | $n = 50$ | 0.080 | 0.057 [16] | 0.072 | **0.046** | **0.045** |
| | (s.e.) | (0.003) | (0.002) | (0.002) | (0.002) | (0.002) |
| | $n = 100$ | 0.047 | 0.035 [12] | 0.040 | **0.024** | **0.023** |
| | (s.e.) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| $c = 1$ | $n = 50$ | 0.156 | 0.114 [30] | **0.082** | 0.083 | **0.081** |
| | (s.e.) | (0.004) | (0.003) | (0.003) | (0.003) | (0.003) |
| | $n = 100$ | 0.099 | 0.079 [35] | **0.045** | 0.062 | **0.060** |
| | (s.e.) | (0.002) | (0.002) | (0.001) | (0.002) | (0.002) |
| Plateau function | | | | | | |
| $c = 0.5$ | $n = 50$ | 0.068 | **0.051** [3] | 0.077 | **0.055** | 0.059 |
| | (s.e.) | (0.003) | (0.002) | (0.002) | (0.002) | (0.002) |
| | $n = 100$ | 0.036 | **0.029** [2] | 0.046 | **0.029** | 0.031 |
| | (s.e.) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| $c = 1$ | $n = 50$ | 0.093 | **0.075** [0] | 0.117 | **0.074** | 0.076 |
| | (s.e.) | (0.003) | (0.002) | (0.003) | (0.002) | (0.003) |
| | $n = 100$ | 0.049 | **0.040** [0] | 0.069 | **0.039** | 0.041 |
| | (s.e.) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) |

function, monotonicity-restricted MGCV is the best performer, but the performance of GMBBoost is very similar. It is obvious that using PAVA is not a good idea when the true underlying function is smooth. Note that the differences between the stopping criteria are marginal. For the step function, BIC does a little better, whereas in the plateau example, AIC has the advantage.

To assess the performance of the GMBBoost extension to the GAM framework, we conduct simulation studies from 2 settings with multiple covariates. We investigate again a Poisson model with logarithmic link, where the linear predictor is given by

$$\eta_i = c\left(\alpha_0 + \sum_{j=1}^{p} f_j(x_{ij})\right) \tag{3.2}$$

and $c$ relates to the strength of the signal. First, $p = 2$ and 2 monotonic increasing functions,

$$f_1(x) = (x - 0.5)^7 + 2(x - 0.3)^5 + (x - 0.7)^3 \quad \text{(polynomial)}$$

and

$$f_2(x) = 1/\{1 + \exp[-50(x - 0.5)]\} \quad \text{(sigmoidal),}$$

are considered ($\alpha_0 = 0.5$). The covariates $x_{i1}$ and $x_{i2}$ are drawn independently from a $U[0, 1]$-distribution. We compare GMBBoost with both smooth components under a monotonicity restriction to the additive extension of MGCV based on cubic regression splines, where the multiple smoothing parameters are selected by UBRE based on Newton's method in multidimensions (see, e.g. Gu and Wahba, 1991, and Wood, 2000). As above, for each smooth component, B-splines of degree 3 with 20 equidistant knots within the domain of the data and a ridge parameter of $\lambda = 300$ are used. We use a dimension of $k = 22$ in each component for MGCV. Table 2 shows the results of the mean AKL over $S = 250$ for 2 different levels of the signal strength and sample sizes of $n = 100$ and 200. In all settings, GMBBoost yields considerably better estimates than the unconstrained GAM. The differences between AIC- and BIC-stopped GMBBoost are again small, with no clear preference for either criterion.

Finally, we consider a setting in higher dimensions, where only some of the components are under monotonicity constraints. A Poisson model with a log-link and a linear predictor as given in (3.2) is considered, where $p = 5$ and the last 3 functions are assumed to be monotonic. See Figure 1 for the shape of the functions ($\alpha_0 = 0$). The algebraic expressions of these functions are given in Appendix A.2. The covariates are drawn from a $\mathcal{N}_5(\mathbf{0}, \mathbf{\Sigma})$-distribution, where a compound symmetry correlation structure of the design is obtained by using $\mathbf{\Sigma} = \rho\mathbf{1}\mathbf{1}' + (1 - \rho)\mathbf{I}$, we chose $\rho = 0.4$. In order to keep the strength of the signal comparable to the example above, rather small values of $c = 0.2$ and 0.3 are chosen.

In this case, GMBBoost selects between 2 different types of weak learners: grouped basis functions with restrictions and P-splines (see Section 2.3). Simulations suggest that different ridge parameters should be used for these 2 types of learners. In the examples presented above, we observed that a ridge

Table 2. *AKL error over $S = 250$ simulated data sets for 2-dimensional Poisson regression, with corresponding standard errors. The number of instances where no fit could be obtained is given in brackets*

| | $c = 1$ | | | $c = 1.5$ | | |
|---|---|---|---|---|---|---|
| | MGCV | GMB (AIC) | GMB (BIC) | MGCV | GMB (AIC) | GMB (BIC) |
| $n = 100$ | 0.058 [3] | 0.033 | 0.032 | 0.070 [0] | 0.044 | 0.045 |
| (s.e.) | (0.002) | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) |
| $n = 200$ | 0.032 [0] | 0.020 | 0.021 | 0.043 [0] | 0.030 | 0.031 |
| (s.e.) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |

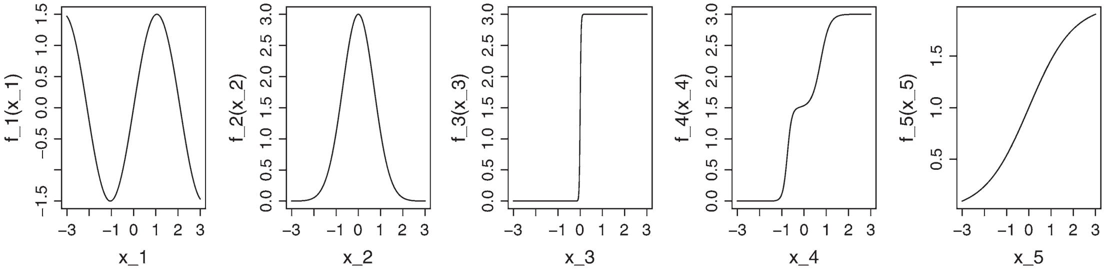

Fig. 1. Functions $f_s(\cdot), s = 1, \ldots, 5$, used for the simulation in higher dimensions. The last 3 functions are monotonic, the first 2 are not.

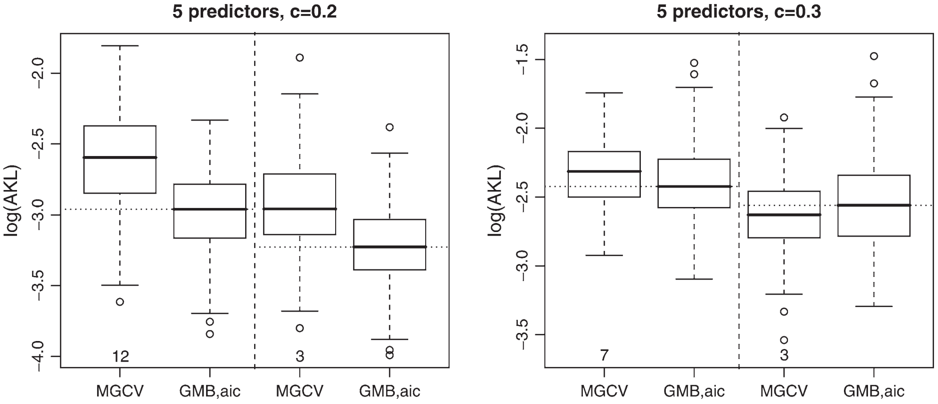

Fig. 2. Boxplots of log(AKL) for different fitting methods for the model with 5 predictors with different noise levels. In each panel, the results of sample sizes for $n = 200$ (left) and $n = 300$ (right) are given. The number of instances where no MGCV fit could be obtained is given at the bottom of the corresponding boxplots.

parameter of $\lambda = 300$ for the grouped B-splines is large enough to yield an appropriate fit. We also tried larger values which required a higher number of iterations, but only with marginally differing results. Hence, $\lambda = 300$ is also applied here. However, we allow for more flexibility in the choice of the P-spline penalty parameter $\lambda_P$. In each setting, GMBBoost cycles for 3 different values of $\lambda_P$, and the combination of $\lambda_P$ and number of iterations $l$ is chosen that minimizes AIC or BIC. A maximum number of $L = 120$ boosting iterations for each cycle suffices in most cases, which makes the procedure computationally feasible. Since small values of $\lambda_P$ may lead to premature stopping, we consider 3 different values of $\lambda_P \in \{1000, 3000, 5000\}$. The choice of basis functions and knots is the same as in the example with $p = 2$. We again compare GMBBoost to MGCV which uses the same settings as before.

In Figure 2, the logarithm of the AKL is given for the various settings. It is seen that GMBBoost outperforms MGCV in most considered settings, with a distinct dominance in the higher noise case $c = 0.2$ (left panel). Interestingly, BIC seems to stop boosting prematurely in the higher dimensional case, yielding an inferior performance compared to AIC-stopped GMBBoost. Thus, we recommend to use the AIC criterion in problems where also nonrestricted smooth components have to be estimated.

## 4. AIR POLLUTION IN SÃO PAULO

In the following, the air pollution data mentioned in Section 1 are investigated more closely. The objective is to evaluate the association between mortality for respiratory causes and the concentration of $SO_2$, CO,

$PM_{10}$, and $O_3$. Previous analyses of this data set (see, e.g. Conceição *and others*, 2001, or Singer *and others*, 2002) focus on the population of children under 5. In the analysis presented here, we consider the risk group of elderly people (aged 65 years or older; see also Saldiva *and others*, 1995). More detailed information about the origin of the data and the preprocessing steps is given in Appendix A.3. The response variable is the number of daily deaths of elderly people attributed to respiratory causes in the city of São Paulo. The sample size is $n = 1351$. 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 (1994b). As the dependent variable consists of count data, we use a Poisson model along with the log-link and consider a core model similar to the model of Singer *and others* (2002),

$$\begin{aligned}\eta_{\text{core}} = \log[E(\text{resp. deaths})] = \alpha_0 + f_1(\text{time}) + \alpha_{01} \cdot \text{temp} + \alpha_{02} \cdot \text{humidity} \\ + \alpha_{03} \cdot \text{Monday} + \cdots + \alpha_{08} \cdot \text{Saturday} \\ + \alpha_{09} \cdot \text{nonresp. deaths}. \end{aligned} \tag{4.1}$$

The model includes a nonspecified function of time (in days) to control for long-term seasonality. We also tried models where temperature (daily minimum, 2-day lagged) and humidity were considered as smooth functions. Since we found no evidence of nonlinearity, we decided to include these covariates as linear terms. In addition, day-of-week dummies are included to control for short-term seasonality and the number of deaths by nonrespiratory 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 focus exclusively on the concentration of $SO_2$, given in daily mean values of $\mu g/m^3$ (2-day lagged), considering the predictor

$$\eta = \eta_{\text{core}} + f_2(SO_2), \tag{4.2}$$

where $f_2(\cdot)$ is monotonic increasing (for a discussion see concluding remarks). To account for that assumption, model (4.2) has been fitted by the boosting procedure described in Section 2 (GMBBoost), where $f_2$ was estimated under the monotonicity constraint. We used a B-spline basis of degree 3 with $\tilde{m} = 20$ equidistant interior knots for each of the smooth component and concentrate on AIC-stopped GMBBoost. When using the same values for the penalty parameters $\lambda$ and $\lambda_P$ as in the simulation study, no distinct minimal AIC-value was attained after $L = 400$ boosting iterations. To keep computation feasible, we decided to use considerably lower values of $\lambda = 30$ and $\lambda_P \in \{30, 50, 100\}$. Then, AIC-stopped GMBBoost chose $\lambda_P = 30$ at an optimal number of $l_{\text{opt}} = 71$ iterations. It should be noted that the results for the other values of $\lambda_P$ are very similar.

The fixed effects were reestimated in each iteration. For comparison, we also fitted a GAM by the use of the R package mgcv with the same adjustments as in the simulations.

Figure 3 shows the estimated curves $\hat{f}_1$ and $\hat{f}_2$ for the various fitting procedures. It is seen from $\hat{f}_1$ (left panel) that a strong seasonal pattern in mortality is evident for both fitting methods. The more interesting result is the difference in the fits for the concentration of $SO_2$ (right panel). The MGCV fit shows a rather wiggly curve up to a concentration of 35 $\mu g/m^3$. Interestingly, the fitted curve decreases for very small concentrations, which is counterintuitive. In contrast, GMBBoost shows different behavior. Since monotonicity is assumed for this component, one obtains a fit that increases first slowly (up to about 40 $\mu g/m^3$) and then steeply (up to 60 $\mu g/m^3$), where it remains constant on a rather high level of mortality for high pollutant concentrations.

In Table 3, the parameter estimates for the fixed effects, controlling for temperature, humidity, long-term seasonality, and nonrespiratory deaths, are given for the different fitting methods, along with the corresponding standard errors. It is seen that the estimates are rather stable across fitting procedures. We also give the deviance and the effective degrees of freedom for both models. It is seen that the MGCV fit is closer to the data, whereas GMBBoost yields a sparser model due to the restriction. Note that for

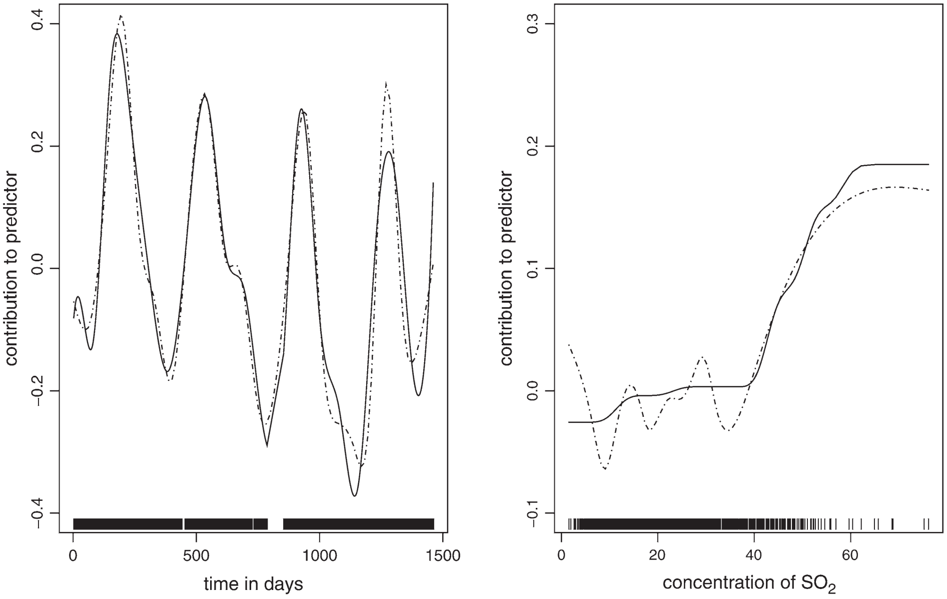

Fig. 3. Core model + $f_2(\mathrm{SO}_2)$, estimated curves for the smooth components, GMBBoost with monotonic fitting of $f_2(\mathrm{SO}_2)$, AIC-stopped (solid), and MGCV (dash-dotted). Data points are given as rug at the foot of each panel.

Table 3. *Core model + $f_2(SO_2)$, estimates of fixed coefficients for MGCV and AIC-stopped GMBBoost, along with corresponding standard errors*

| | MGCV | s.e. | GMBBoost, AIC | s.e. |
|---|---|---|---|---|
| Intercept | 2.6337 | 0.1184 | 2.5383 | 0.1505 |
| Minimum temperature | −0.0015 | 0.0038 | −0.0014 | 0.0038 |
| Humidity | −0.0033 | 0.0009 | −0.0028 | 0.0009 |
| Monday | 0.0316 | 0.0292 | 0.0250 | 0.0294 |
| Tuesday | −0.0397 | 0.0310 | −0.0449 | 0.0295 |
| Wednesday | −0.0250 | 0.0289 | −0.0290 | 0.0286 |
| Thursday | −0.0481 | 0.0289 | −0.0503 | 0.0284 |
| Friday | −0.0599 | 0.0290 | −0.0632 | 0.0283 |
| Saturday | −0.0331 | 0.0288 | −0.0359 | 0.0285 |
| Nonrespiratory deaths | 0.0029 | 0.0009 | 0.0038 | 0.0009 |
| Deviance | 1386.8935 | | 1430.9318 | |
| edf | 40.4720 | | 32.6385 | |
| $\hat{\phi}$ | 1.0523 | | 1.0840 | |

GMBBoost, the effective degrees of freedom have to be interpreted with caution, since they depend on the approximation of the hat matrix.

The editor pointed out with good reason that overdispersion may be a problem in Poisson models. We therefore estimated the scale parameter,

$$\hat{\phi} = \frac{1}{n - \mathrm{edf}} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i},$$

for both models, where edf denotes the effective degrees of freedom. From Table 3, it is seen that $\hat{\phi}$ is in both cases fairly close to one, indicating that overdispersion should not be an issue in this application. In addition, we computed the partial serial correlation of residuals. The inclusion of days of the week as covariates had the effect that partial serial correlation is rather weak. The maximal value was 0.062. For significance level 0.05 for single lags, 2 lags out of 30 (1 and 13) were just above the critical value. When adjusting for multiple tests, serial correlation seems hardly relevant. Similar results were reported by Singer *and others* (2002).

Figure 4 shows the curves fitted by AIC-stopped GMBBoost and approximate 0.95 pointwise confidence bands as derived in Appendix A.1. Although confidence intervals are rather large for a $SO_2$ concentration above 40 μg/m$^3$, the confidence intervals show that the increase should be taken seriously.

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 *and others*, 2002, or Einbeck *and others*, 2004). Let $SO_2(i)$ be the recorded concentration in observation $i$, $i = 1, \dots, n$, and $SO_2(\text{min})$ the minimum concentration recorded, then the relative risk of death is defined by

$$\begin{aligned}\mathrm{RR}(i) &= \frac{E(\text{respiratory death}|SO_2(i))}{E(\text{respiratory death}|SO_2(\text{min}))} = \frac{\exp[\eta_{\text{core},i} + f_2(SO_2(i))]}{\exp[\eta_{\text{core},i} + f_2(SO_2(\text{min}))]}\\ &= \exp[f_2(SO_2(i)) - f_2(SO_2(\text{min}))].\end{aligned}$$

In Figure 5, the estimated relative risk curve is given for 2 fitting methods. It is obvious that the non-monotonic fit yields unreliable results with a fit which indicates that the relative risk of death is below 1 up to 45 μg/m$^3$. In contrast, the GMBBoost fit shows monotonic increase over the whole risk curve. For concentrations higher than 60 μg/m$^3$, the risk seems to remain on a fairly high level for the AIC-stopped boosting.

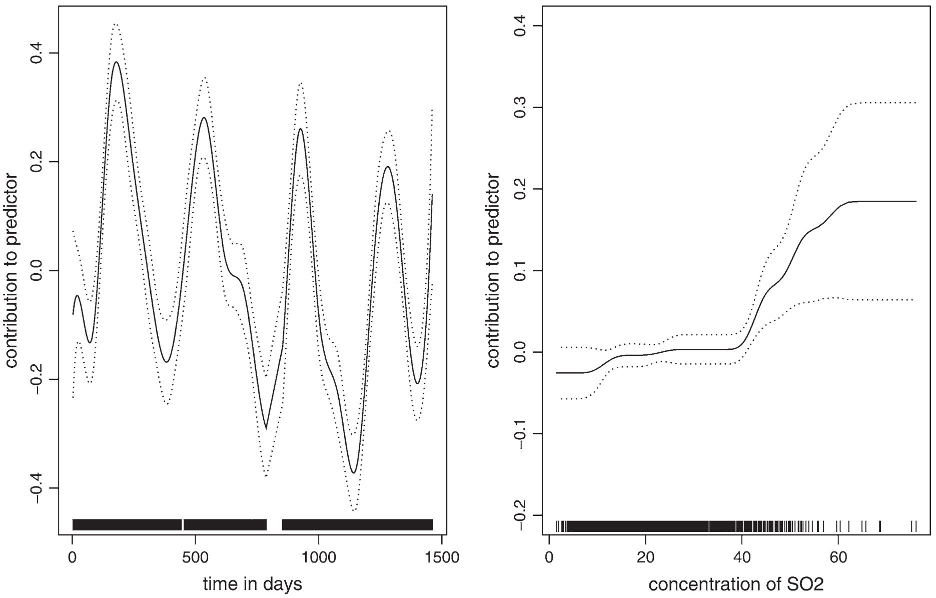

Fig. 4. Core model + $f_2(SO_2)$, AIC-stopped GMBBoost with monotonic fitting of $f_2(SO_2)$ along with the approximate confidence intervals. Data points are given as rug at the foot of each panel.

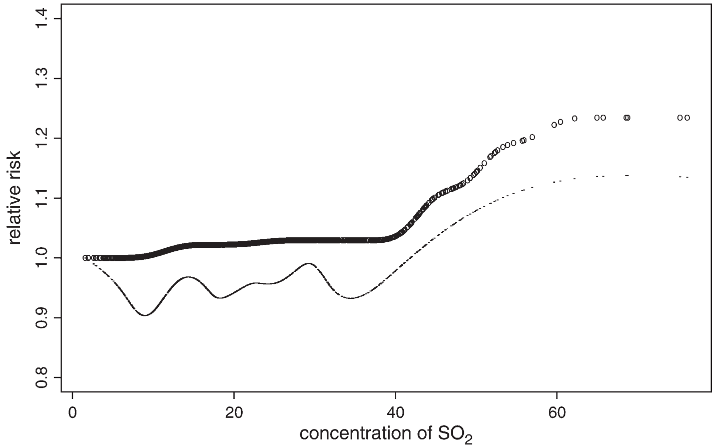

Fig. 5. Relative risk curves versus $SO_2$ concentration for MGCV (·) and AIC-stopped GMBBoost (○).

## 5. Concluding remarks

A procedure is proposed that incorporates monotonicity constraints for one or more components within a GAM. By using monotonicity, the procedure yields stable estimates which in contrast to GAM fitting avoids overfitting and questionable wiggly estimates. Einbeck *and others* (2004) try to find stable estimates by downweighting observations. 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 the case of higher dimensional covariates is hard to tackle—is avoided by boosting techniques. The only crucial tuning parameter is the number of boosting iterations, which is chosen by the AIC or the BIC criterion. Moreover, the procedure is a strong competitor to alternative approaches.

The results provided by the proposed methodology go with other findings regarding the relationship between air pollution and mortality in the São Paulo area. In an earlier study, Saldiva *and others* (1995) investigated all causes of mortality of elderly people. They found a significant effect of $SO_2$ by using a parametric Poisson model, controlling for seasonality by dummy variables for months and years. Later, Pereira *and others* (1998) investigated the association between air pollution and mortality of fetuses with a parametric Poisson model, reported a weakly increasing risk of death for $SO_2$, and derived a dose–response relationship between the pollutant and intrauterine mortality. In the study by Conceição *and others* (2001), the respiratory mortality of children under 5 was investigated by using GAM techniques and the same database as this paper. They found an increase in relative risk of death of about a 25% on the most polluted days (compared to the least polluted days) for $SO_2$. This is comparable to the relative risk of death that we observed for elderly people using the GMBBoost procedure, see Figure 5. Braga *and others* (2001) explored hospital admissions due to respiratory causes of children and adolescents under 20. They observed a significant effect of $SO_2$ on infants under 2 but no significance in other age groups.

To conclude, previous investigations of adverse health effects of $SO_2$ in São Paulo are supported by our approach. Besides the known effects on children's health (Conceição *and others* (2001)), $SO_2$ seems also to affect the health of elderly people. A particular characteristic of the effects of the pollutant $SO_2$ in São Paulo is that most of the pollution in this city is caused by auto engine exhaust gases. $SO_2$ (as well as CO) may be seen as good proxies for automotive emissions.

A referee raised the problem that there is not necessarily an increasing relationship between high air pollution concentration and mortality. When ambient air pollution is at a very high level, people may adjust their behavior, for example they may stay inside on particularly polluted days (see Bresnahan *and others*, 1997, or Neidell, 2002, for a detailed discussion). Nevertheless, there is empirical evidence that high pollutant concentrations result in an increasing mortality. Thus, we think that it is reasonable to consider ambient air pollution as a proxy for air pollution exposure.

An implementation in R of the approach outlined in this paper is available at http://www.statistik.lmu.de/institut/lehrstuhl/semsto/Software/software.htm.

## ACKNOWLEDGMENTS

We gratefully acknowledge support from the Deutsche Forschungsgemeinschaft (SFB 386, “Statistical Analysis of Discrete Structures”). We thank the “Laboratorio de Poluição Atmosférica Experimental, Faculdade de Medicina, Universidade de São Paulo, Brazil”, Jochen Einbeck, and Julio M. Singer for letting us use the data on air pollution in São Paulo. We also thank Harald Binder and 2 unknown referees for constructive comments. *Conflict of Interest:* None declared.

## APPENDIX

### A.1 *Standard deviations*

In order to obtain standard deviations for function estimates, we suggest starting from the approximate hat matrix given in (2.9). Consider the model from (2.10), where components $1, \ldots, \upsilon$ are estimated under the monotonicity constraint and components $\upsilon + 1, \ldots, p$ are not; the linear predictor after $l$ boosting iterations is given by

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

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

$$\mathbf{B}^{(s)}\hat{\boldsymbol{\alpha}}_s^{(k)} = \begin{cases} \mathbf{B}^{(s)}\hat{\boldsymbol{\alpha}}_s^{(k-1)}, & s \neq s_k, \\ \mathbf{B}^{(s)}\hat{\boldsymbol{\alpha}}_s^{(k-1)} + \mathbf{S}_k(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k-1)}), & s = s_k, \end{cases} \tag{A.2}$$

where, according to (2.10) and (2.12),

$$\mathbf{S}_k = \begin{cases} \mathbf{B}_{(r_k)}^{(s_k)}\big(\mathbf{B}_{(r_k)}^{(s_k)'}\mathbf{W}_k\mathbf{B}_{(r_k)}^{(s_k)} + \lambda\boldsymbol{\Lambda}\big)^{-1}\mathbf{B}_{(r_k)}^{(s_k)'}\mathbf{W}_k\mathbf{D}_k{}^{-1}, & s_k \leqslant \upsilon, \\ \mathbf{B}^{(s_k)}(\mathbf{B}^{(s_k)'}\mathbf{W}_k\mathbf{B}^{(s_k)} + \lambda_{\mathrm{P}}\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2)^{-1}\mathbf{B}^{(s_k)'}\mathbf{W}_k\mathbf{D}_k{}^{-1}, & s_k > \upsilon. \end{cases} \tag{A.3}$$

From (A.3), it becomes apparent that the type of the update of the chosen smooth component depends on the presence or absence of a monotonicity restriction. Using the indicator function $I(\cdot)$, (A.2) can be written in the closed form

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

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

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

and hence, in a recursive fashion,

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

where

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

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

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

where $\widehat{\text{cov}}(\mathbf{y}) = \text{diag}\{\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2\}$, with $\hat{\sigma}_i^2 = \widehat{\text{var}}(y_i)$.

Standard deviations for the fixed effects may be obtained in a similar way. The vector of fixed parameters after $l$ boosting iterations is given by

$$\hat{\boldsymbol{\alpha}}_0^{(l)} = \hat{\boldsymbol{\alpha}}_0^{(l-1)} + (\mathbf{Z}'\tilde{\mathbf{W}}_l\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{W}}_l\tilde{\mathbf{D}}_l^{-1}(\mathbf{y} - \tilde{\boldsymbol{\mu}}^{(l)}). \tag{A.4}$$

Let $\mathbf{T}_k = (\mathbf{Z}'\tilde{\mathbf{W}}_k\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{W}}_k\tilde{\mathbf{D}}_k^{-1}$ and use the approximation of the hat matrix, $\tilde{\boldsymbol{\mu}}^{(k)} \approx \tilde{\mathbf{H}}_k\mathbf{y}$, where $\tilde{\mathbf{H}}_k = \mathbf{I} - (\mathbf{I} - \mathbf{M}_k)\big[\prod_{j=1}^{k-1}(\mathbf{I} - \tilde{\mathbf{M}}_{k-j})(\mathbf{I} - \mathbf{M}_{k-j})\big](\mathbf{I} - \mathbf{M}_0)$, then (A.4) may be expressed in a recursive way as

$$\hat{\boldsymbol{\alpha}}_0^{(l)} \approx \sum_{k=0}^{l}\mathbf{T}_k(\mathbf{I} - \tilde{\mathbf{H}}_k)\mathbf{y},$$

where $\mathbf{T}_0 = \frac{1}{n}(\mathbf{1}_n, \mathbf{0}_{(n\times(u-1))})'$ and $\tilde{\mathbf{H}}_0 = \mathbf{0}_{(n\times n)}$. With $\mathbf{U}_l = \sum_{k=0}^{l}\mathbf{T}_k(\mathbf{I} - \tilde{\mathbf{H}}_k)$, it immediately follows that

$$\widehat{\text{cov}}(\hat{\boldsymbol{\alpha}}_0^{(l)}) \approx \widehat{\text{cov}}(\mathbf{U}_l\mathbf{y}) = \mathbf{U}_l\widehat{\text{cov}}(\mathbf{y})\mathbf{U}_l'.$$

### A.2 *Algebraic expressions of smooth functions*

The smooth functions used in the simulation study of the 5-dimensional additive Poisson model (see Figure 1) are given in algebraic terms by

$$f_1(x_1) = 1.5\sin(1.5x_1),$$
$$f_2(x_2) = 3\exp(-x_2^2),$$
$$f_3(x_3) = 3/[1 + \exp(-50x_3)],$$
$$f_4(x_4) = 1.5/\{1 + \exp[-10(x_4 + 0.75)]\} + 1.5/\{1 + \exp[-5(x_4 - 0.75)]\},$$
$$f_5(x_5) = 2/[1 + \exp(-x_5)].$$

### A.3 *Preprocessing of the data set*

The data stem from a study conducted between January 1994 and December 1997. The original database can be found on the web at

http://www.ime.usp.br/˜jmsinger/Polatm9497.zip (the data file is called master.xls). The concentrations of several air pollutants were recorded daily by 13 monitoring stations of the São Paulo state air pollution controlling agency. For $SO_2$, 24-h mean values (in $\mu g/m^3$) are given (variables SO21 to SO221). We took the mean over all monitoring station where records are available to get an overall variable for $SO_2$ concentration. The daily minimum temperature (TEMPMIN) and relative humidity (HUMIDMED) are provided by the University of São Paulo. We follow Conceição *and others* (2001) and consider the measurements taken 2 days before as influential. This refers to the measurements of the $SO_2$ concentration and the minimum temperature, as well as the measurement of the concurrent day is used for relative humidity. Daily records of mortality were recorded from the municipal mortality information improvement program. For several age groups, the total number of deaths is given, as well as a more detailed partition into deaths due to respiratory diseases, cardiovascular, and other causes. In the present analysis, the respiratory mortality of population at an age of 65 years or older is investigated. We use the daily number of deaths in this age group due to respiratory diseases (RES65) as response variable. The mortality due to nonrespiratory causes, which is used as a covariate, contains the deaths due to cardiovascular (CAR65) and other (OTH65) causes. For a more detailed description of the classification of mortality, we refer to Conceição *and others* (2001). The preprocessing results in a data set of $n = 1351$ complete observations.

## References

Braga, A. L. F., Saldiva, P. H. N., Pereira, L. A. A., Menezes, J. J. C., Conceição, G. M. S., Lin, C. A., Zanobetti, A., Schwartz, J. and Dockery, D. W. (2001). Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil. *Pediatric Pulmonology* **31**, 106–113.

Bresnahan, B., Dickie, M. and Gerking, S. (1997). Averting behavior and urban air pollution. *Land Economics* **73**, 340–357.

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*. Munich, Germany: LMU München.

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

Bühlmann, P. and Yu, B. (2003). Boosting with the $L_2$-loss: regression and classification. *Journal of the American Statistical Association* **98**, 324–339.

Bühlmann, P. and Yu, B. (2006). Sparse boosting. *Journal of Machine Learning Research* **7**, 1001–1024.

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.

Dilleen, M., Heimann, G. and Hirsch, I. (2003). Non-parametric estimators of a monotonic dose-response curve and bootstrap confidence intervals. *Statistics in Medicine* **22**, 869–882.

Ducharme, G. R. and Fontez, B. (2004). A smooth test of goodness-of-fit for growth curves and monotonic nonlinear regression models. *Biometrics* **60**, 977–986.

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.

GU, C. AND WAHBA, G. (1991). Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method. *SIAM Journal of Scientific and Statistical Computing* **12**, 383–398.

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.

LEE, C.-I. C. (1996). On estimation for monotone dose-response curves. *Journal of the American Statistical Association* **91**, 1110–1119.

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

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

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

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

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

NEIDELL, M. (2002). Air pollution, health and socio-economic status: the effect of outdoor air quality on childhood asthma. *Journal of Health Economics* **23**, 1209–1236.

PEREIRA, L. A. A., LOOMIS, D., CONCEIÇÃO, G. M. S., BRAGA, A. L. F., ARCAS, R. M., KISHI, H. S., SINGER, J. M., BÖHM, G. M. AND SALDIVA, P. H. N. (1998). Association between air pollution and intrauterine mortality in São Paulo, Brazil. *Environmental Health Perspectives* **106**, 325–329.

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

RAMSAY, J. O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society B* **60**, 365–375.

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

SALDIVA, P. H. N., POPE, III, C. A., SCHWARTZ, J., DOCKERY, D. W., LICHTENFELS, A. J. F. C., SALGE, J. M., BARONE, I. A. AND BÖHM, G. M. (1995). Air pollution and mortality in elderly people: a time-series study in São Paulo, Brazil. *Archives of Environmental Health* **50**, 159–163.

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

SCHWARTZ, J. (1994a). Air pollution and daily mortality: a review and meta analysis. *Environmental Research* **64**, 36–52.

SCHWARTZ, J. (1994b). Nonparametric smoothing in the analysis of air pollution and respiratory illness. *Canadian Journal of Statistics* **22**, 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). Association between atmospheric pollution and mortality in São Paulo, Brazil: regression models and analysis strategy. In: Dodge, Y. (ed.), *Statistical Data Analysis Based on the L1 Norm and Related Methods*. Berlin: Birkhuser. pp 439–450.

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

TURLACH, B. A. (2005). Shape constrained smoothing using smoothing splines. *Computational Statistics* **20**, 81–103.

TUTZ, G. AND BINDER, H. (2006). Generalized additive modelling with implicit variable selection by likelihood based boosting. *Biometrics* (in press).

TUTZ, G. AND LEITENSTORFER, F. (2006). Generalized smooth monotonic regression in additive modeling. *Journal of Computational and Graphical Statistics* (in press).

WOOD, S. N. (1994). Monotonic smoothing splines fitted by cross validation. *SIAM Journal on Scientific Computing* **15**, 1126–1133.

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

WOOD, S. N. (2001). mgcv: GAMs and generalized ridge regression for R. *R News* **1**, 20–25.

WOOD, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society B* **65**, 95–114.

ZHANG, J. T. (2004). A simple and efficient monotone smoother using smoothing splines. *Journal of Nonparametric Statistics* **16**, 779–796.

[*Received June 2, 2005; first revision March 21, 2006; second revision July 27, 2006; accepted for publication October 20, 2006* ]