Increasing the usefulness of additive spline models by knot removal

Harald Binder & Willi Sauerbrei

Universität Freiburg i. Br.

Nr. 99 October 2007

Zentrum für Datenanalyse und Modellbildung
Universität Freiburg
Eckerstraße 1
D–79104 Freiburg im Breisgau

und

Institut für Medizinische Biometrie und Medizinische Informatik
Universitätsklinikum Freiburg
Stefan-Meier-Straße 26
D–79104 Freiburg im Breisgau

binderh@fdm.uni-freiburg.de
wfs@imbi.uni-freiburg.de

Abstract

Modern techniques for fitting generalized additive models mostly rely on basis expansions of covariates using a large number of basis functions and penalized estimation of parameters. We exemplarily use a mixed model approach to fit a model for children's lung function that allows for non-linear influence of several covariates available in a substantial data set. While the resulting model is expected to have good prediction performance, its handling beyond simple visual presentation is problematic. It is shown how the number basis functions of the underlying B-spline representation can be reduced by knot removal techniques without refitting, while preserving the shape of the fitted functions. We extend the condition for exact knot removal towards approximate knot removal by incorporating the covariance matrix of the initial parameter estimates, resulting in considerable simplification of the model. Covariance matrices for the transformed parameter estimates are provided. It is demonstrated that enforcing the knot removal condition during estimation leads to the difference penalties employed in the P-spline approach for estimation of B-spline coefficients and therefore provides a further justification for this type of penalty. A final transform to a truncated power basis provides a simple equation for the model. This increases transportability, while retaining properties of the initial fit such as good prediction performance.

Key words: B-splines; Difference penalty; Generalized additive models; Knot-removal; Transportability

1 Introduction

In medical research there are many diagnostic and prognostic studies where for each observation a considerable number of continuous covariates is available and where for each of them identification of the functional form of its influence on the response is wanted. Generalized additive models (Hastie and Tibshirani, 1990) provide a convenient framework where continuous covariates are modeled to have additive influence, specified by unknown functions on an exponential family response. This includes models for different types of response, such as continuous, binary, or Poisson responses.

Modern techniques for fitting such models, e.g. via simultaneous estimation of all functions by penalized likelihood (Marx and Eilers, 1998; Wood, 2004, 2006) or via a mixed model representation (see Ruppert et al., 2003, for example), have in common that they use basis expansions of covariates. A set of knots covering the domain of the covariate is employed to obtain an expanded set of variables that allows for modeling of local features. While this may result in good prediction performance when the smoothness of the fitted functions is properly regularized (Binder and Tutz, 2006), the resulting models are difficult to transport beyond simple visual presentation. For example, a suggested number of 20 or more basis functions per covariate (Ruppert, 2002) implies that already for univariate models no simple model equation is available. Obviously, a simple model equation featuring only a few parameters would facilitate communication and transportability of the model (e.g. doctors could easily calculate the model's prediction for a patient) and thereby increase its clinical usefulness (Wyatt and Altman, 1995).

In the following we present a technique for knot removal that reduces the number of basis functions while preserving the shape of the fitted functions and thereby also prediction performance. Especially when a fitted function has simple structure (e.g. when the plot looks like a global polynomial) a substantial number of knots can be eliminated. The coefficients for the resulting smaller set of new basis functions is obtained without refitting by a linear transform and therefore also transformed covariance matrices are available. The technique will first be introduced for univariate models, but as there is no refitting required it is then easily transferred to the multivariate setting.

There are many univariate approaches for forward knot selection (see Zhou and Shen, 2001; Miyata and Shen, 2003; Mao and Zhao, 2003; Gervini, 2006, for example) and also some prominent multivariate approaches such as TURBO (Friedman and Silverman, 1989), MARS (Friedman, 1991), POLYMARS/POLYCLASS (Stone et al., 1997; Kooperberg et al., 1997), or the approaches of Breiman (1993) and Molinari et al. (2004). These are distinctly different from the present knot removal approach, because they employ refitting after every knot selection step. That way they try to simultaneously address the objectives of good prediction performance and of obtaining a parsimonious model. In contrast, the present approach relies on the good prediction performance of existing fitting techniques for generalized additive models seen for example in Binder and Tutz (2006). Model simplification is only attempted afterwards in a shape-preserving manner without refitting. This approach is also supported by results from preliminary experiments where refitting after knot removal did even result in a slight decrease of prediction performance on new data.

Section 2 gives a formal description of the generalized additive modeling framework and reviews the method for fitting such models that will be used as a basis for knot removal, including the B-spline basis and penalized estimation using the difference penalty from the P-spline approach. Exact knot-removal based on linear transforms will be reviewed in Section 3.1 and an extended version for approximate knot removal together with transformed covariance matrices will be introduced in Section 3.2. This includes a result that motivates the difference penalty of the P-spline approach as an enforcement of knot removal at estimation (Section 3.3). In Section 4 a final transform from the reduced B-spline basis to a truncated power series basis is proposed for obtaining interpretable model equations. Section 5 presents examples where the proposed techniques are applied to models fitted to children's lung function data aiming to obtain simple model equations. Finally, Section 6 gives a discussion and sketches potential extensions.

2 Fitting Generalized Additive Models

2.1 Framework

Given data $(y_i, x_i), i = 1, \ldots, n$ with response y_i and covariate vectors $x_i = (x_{i1}, \ldots, x_{ip})'$, generalized additive models (Hastie and Tibshirani, 1990) assume that the response is from an exponential family, which (among others) allows for modeling of continuous, binary as well as Poisson responses. The structural part, which allows for non-linear influence of the covariates via unknown functions $f_j, j = 1, \ldots, p$, is given by

$$E(y_i|x_i) = \mu_i = g\left(\beta_0 + \sum_{j=1}^{p} f_j(x_{ij})\right)$$

with known response function g and intercept parameter β_0. Modern approaches feature simultaneous estimation of all functions f_j, e.g. by direct maximization of a penalized likelihood (Marx and Eilers, 1998; Wood, 2004) or via a mixed model representation (see Ruppert et al., 2003, for example). They require basis expansions of covariates, e.g. using a B-spline or a truncated power series basis, which will be reviewed in the following.

Note that even if the same type of basis expansion is used for different estimation approaches, probably the resulting models will differ. For example, the mixed model approach is known to result in smoother fitted functions (Binder and Tutz, 2006) than direct maximization of penalized likelihood. If comparison beyond graphical displays is wanted, it is necessary to simplify the model equation for each covariate.

2.2 Truncated Power Series and B-Spline Basis

For one covariate a function f is represented by a combination of M basis functions

$$f(x) = \sum_{k=1}^{M} \delta_k B_k(x) \tag{1}$$

with known basis functions $B_k, k = 1, \ldots, M$ and parameters $\delta_k, k = 1, \ldots, M$ that have to be estimated. A popular choice for the basis functions, advocated for example by Ruppert et al. (2003), is the truncated power series basis. Let the domain of the covariate x be divided into s intervals by an arbitrary knot sequence $\xi_1 < \ldots < \xi_{s+1}$, where we say that knots $\xi_2, \ldots, \xi_s$ are "within" the domain of the covariate. With a truncated power basis of degree q the function is given by

$$f(x) = \sum_{l=0}^{q} \delta_{l+1} x^l + \sum_{l=2}^{s} \delta_{l+q}(x - \xi_l)_+^q$$

where $()_+$ takes the positive part of its argument. This has the natural interpretation of a polynomial of degree q, i.e. a global function with $q+1$ components, with (local) modifications to the right of each knot given by $s-1$ truncated power functions. For example, when a smooth function for the effect of a variable x is fitted and there is only a single knot at $x=20$, then the explanation given could be like "the general shape is ..., but changes with $x=20$ ...". The final number of knots within the domain of the covariate, $s-1$, will also be useful to characterize the complexity of the representation of a function after knot removal. For $s=1$ only one interval remains, which means that a global polynomial representation could be obtained.

Another popular choice, which is numerically more stable compared to the truncated power series, is the B-spline basis (de Boor, 2001). For B-splines of degree q the basis functions B_k are polynomial pieces of degree q, which are non-zero on a domain spanned by $q+2$ knots. The position of each piece is characterized by an indexing knot and we take the leftmost knot, the knot at which the polynomial piece starts to become non-zero. As each interval is covered by $q+1$ of these polynomial pieces, compared to the truncated power series basis q knots $\xi_{-q+1},\ldots,\xi_0$ to the left of ξ_1 and q knots $\xi_{s+2},\ldots,\xi_{s+q+1}$ to the right of ξ_{s+1} have to be inserted, resulting in $s+2q+1$ knots for $M=s+q$ basis functions. In contrast to the truncated power series basis there is no distinction between global and local components.

The B-spline basis functions B_k can be defined in terms of truncated power functions $(\xi_l - x)_+^q$, i.e. a set of functions of x indexed by the knots ξ_l (de Boor, 2001, p. 87). Similar to Eilers and Marx (2004) we will deviate slightly by giving the definition based on $(x-\xi_l)_+^q$ (requiring an extra term $(-1)^q$), because this will provide for an easier link to the truncated power series basis. A B-spline basis of degree q can then be defined as

$$B_k(x) = (\xi_{k+q+1} - \xi_k)(-1)^q \left([\xi_k,\ldots,\xi_{k+q+1}](x-\cdot)_+^q\right) \tag{2}$$

where $(x-\cdot)_+^q$ is a "placeholder" notation indicating that x is kept fixed and that $(x-\xi_l)_+^q$ is used as a function of the knots ξ_l alone for determining the divided differences $[\xi_k,\ldots,\xi_{k+q+1}]$ (see de Boor, 2001, for more details).

From (2) it is then seen that B-splines can be defined in terms differences of truncated power functions. This allows for a basis transform from a B-spline basis to a truncated power series basis (see Section 4). It can also be seen that a B-splines basis of degree q corresponds to a truncated power series basis of degree q.

2.3 P-Spline Difference Penalty and Mixed Model Approach

Traditionally only a small number of basis functions for each covariate was incorporated into regression models to avoid overfitting. The corresponding knots were often placed to be equidistant or on quantiles of the data. This results in a certain arbitrariness with respect to the fitted functions because when the number of knots is rather small the exact position may have a strong influence. Alternatively a large number of knots can be used, now combined with penalized estimation. When the number of knots is large enough, neither their exact position nor their exact number seems to matter much when the penalty is chosen adequately to prevent overfitting (Ruppert, 2002). We will use equally spaced knots corresponding to 10 or 20 basis functions in the following.

In their "P-spline" penalty approach, which we will adapt in the following, Eilers and Marx (1996) propose to use a B-spline basis with equally spaced knots together with maximization of a criterion that penalizes for complexity in the form of differences of the parameters. The log-likelihood $l(\delta)$, which classically is maximized for estimation of the parameter vector $\delta^T = (\delta_1, \ldots, \delta_M)$, is augmented by a penalty term to arrive at a penalized criterion

$$l_P(\delta) = l(\delta) - \lambda \sum_{j=k+1}^{M} (\Delta^k \delta_j)^2 \tag{3}$$

where Δ^k denotes the kth order difference and λ is a penalty parameter. The latter determines how smooth the estimated functions will be, i.e. model complexity, and has to be chosen. However, a good choice of the penalty is a difficult issue.

As already pointed out by Speed (1991), penalized estimation of smooth functions can alternatively be performed via a mixed model representation. There the penalty parameters correspond to variance components, i.e. they are determined automatically. Ruppert et al. (2003) provide a comprehensive overview of this approach. In Binder and Tutz (2006) it was seen that it results in parsimonious fits featuring very smooth functions in combination with good prediction performance. As this nicely complements with our aim of obtaining simple model equations, we employ this approach in the following. Specifically, we use the methods described in the appendix of Wood (2004), that allow for incorporating the P-spline difference penalty from (3), together with covariance matrices based on the Bayesian approach described in Wood (2006, p. 318).

To give a theoretical justification for the difference penalty in (3), Eilers and Marx (1996) point out that, using an appropriate differencing order, it can approximate the integrated squared second derivative as a penalty term, which is the standard for smoothing spline estimation (see Green and Silverman, 1994, for example). Note that this is only the case for equally space knots and else weighted differences are required for this connection to hold. A further convenient property of the difference penalty is that for large values of the penalty parameter λ polynomial regression models are obtained (Eilers and Marx, 2004), i.e. the model equation is very simple but still retains the flexibility of a global polynomial when all local structure is smoothed out. Eilers and Marx (2004) also demonstrate that a $q + 1$th order difference penalty for a B-spline basis of degree q corresponds to estimation using a truncated power series basis of degree q together with a ridge penalty (i.e. penalizing the sum of the squared parameter values). Later, when we investigate how knot removal can be enforced at the time of estimation it will be seen that a further justification for the difference penalty is obtained (see Section 3.3).

3 Knot Removal for B-Spline Fits

We investigate a knot removal algorithm that reduces the number of basis functions without refitting while preserving the shape of the fitted functions. First an approach for exact knot removal is reviewed, where the function numerically stays the same. This is subsequently extended towards approximate knot removal where small deviations from the original function are allowed for. Knot removal will be performed separately for each covariate. As there is no refitting, the fitted functions for the other components will not be influenced when knot removal is performed for a specific function. Therefore a univariate strategy for knot removal can easily be extended towards knot removal for multivariate models.

3.1 Exact Knot Removal

Knot removal for B-splines without change of the functional form has been proposed as the reverse of knot insertion (Boehm, 1980; Tiller, 1992). The motivation for knot insertion mainly comes from Computer-Aided Design where new control points have to be added to a surface that is represented by a B-spline basis without changing the shape. Such new control points can then be moved to change the shape of the B-spline in a desired way. In the reverse, knots that have been inserted, but not moved, or knots that fulfill the same conditions as newly inserted knots, can be removed without changing of the shape of the function. Therefore to derive conditions for knot removal, knot insertion has to be examined first.

Given the knots $\xi_{-q+1} < \ldots < \xi_{s+q+1}$ with corresponding B-spline basis functions $B_k, k = 1, \ldots, M$, assume that a new knot $\xi^*, \xi_l < \xi^* < \xi_{l+1}, 1 \leq l < s+1$ is inserted to arrive at knots $\xi^*_{-q+1} < \ldots < \xi^*_{s+q+2}$ with corresponding basis functions $B^*_k(x), k = 1, \ldots, M+1$. This means that a function f in addition to (1) has another unique representation

$$f(x) = \sum_{k=1}^{M+1} \delta^*_k B^*_k(x)$$

in terms of the new knots and basis functions.

Boehm (1980) shows that the parameters $\delta_k, k = 1, \ldots, M$ and $\delta^*_k, k = 1, \ldots, M+1$ of the two representations are related by

$$\delta^*_k = \alpha_{k-q}\delta_k + (1 - \alpha_{k-q})\delta_{k-1} \tag{4}$$

where

$$\alpha_k = \begin{cases} 1 & \text{for } k \leq l-q \\ \frac{\xi^* - \xi^*_k}{\xi^*_{k+q+1} - \xi^*_k} = \frac{\xi^* - \xi_k}{\xi_{k+q} - \xi_k} & \text{for } l-q+1 \leq k \leq l \\ 0 & \text{for } k \geq l+1 \end{cases} .$$

Taking knot removal as the reversal of knot insertion, the parameters after knot removal $\delta_k, k = 1, \ldots, M$ can be calculated recursively from the parameters

$\delta^*_k, k = 1, \ldots, M+1$ before removal (as shown for example by Tiller, 1992) via the relation

$$\delta_k = \begin{cases} \delta^*_k & \text{for } k \leq l \\ (\delta^*_k - (1-\alpha_{k-q})\delta_{k-1})/\alpha_{k-q} & \text{for } l < k < l+q \\ \delta^*_{k+1} & \text{for } k \geq l+q \end{cases} .$$

In matrix form the parameter vector $\delta = (\delta_1, \ldots, \delta_M)^T$ resulting from removal of knot $l+1$ is obtained by

$$\delta = R_{q,l+1}\delta^* \tag{5}$$

with $\delta^* = (\delta^*_1, \ldots, \delta^*_l, \delta^*_{l+2}, \ldots, \delta^*_{M+1})^T$ and

$$R_{q,l+1} = \begin{pmatrix} I_{l-1} & & & \\ & \Xi_{q,l+1} & 0_{q+1} & \\ & & & I_{M-q-l+1} \end{pmatrix}$$

where I_{l-q-1} and $I_{M-q-l+1}$ are identity matrixes of size $l-q-1$ and $M-q-l+1$ respectively, 0_{q+1} is a column vector of zeros of size q, and $\Xi_{q,l+1}$ is a square matrix of size q that has the following structure

$$\Xi_{q,l+1} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & -\frac{1-\alpha_l}{\alpha_l} & \frac{1}{\alpha_l} \end{pmatrix} \cdot \ \cdots \ \cdot \begin{pmatrix} 1 & & & \\ -\frac{1-\alpha_{l-q+2}}{\alpha_{l-q+2}} & \frac{1}{\alpha_{l-q+2}} & & \\ & & 1 & \\ & & & \ddots \end{pmatrix}.$$

Removing a just inserted knot is a rather artificial scenario. In practice there will be an estimated parameter vector $\hat{\delta}^* = (\hat{\delta}^*_1, \ldots, \hat{\delta}^*_M)^T$ corresponding to knot vector $(\xi^*_{-q+1}, \ldots, \xi^*_{s+q+1})^T$. A transformed parameter vector is then obtained by $\hat{\delta}^T = R_{q,l+1}\hat{\delta}^*$ with corresponding new knot vector $(\xi_{-q+1}, \ldots, \xi_{s+q})^T = (\xi^*_{-q+1}, \ldots, \xi^*_l, \xi^*_{l+2}, \ldots, \xi^*_{s+q+1})^T$. Using the matrix formulation (5) a covariance matrix for the transformed estimate $\hat{\delta}$ can be obtained by

$$Cov(\hat{\delta}) = R_{q,l+1} Cov(\hat{\delta}^*) R^T_{q,l+1} \tag{6}$$

where $Cov(\hat{\delta}^*)$ is the covariance matrix of $\hat{\delta}^*$ with the row and column $l+1$ deleted.

For each knot ξ_{l+1} with corresponding parameter $\hat{\delta}^*_{l+1+1}$ it has first to be checked whether it can be removed. The equations in (4) are over-determined and can be used to derive a condition for knot removal. For example for $q = 3$ it follows that

$$\alpha_l \hat{\delta}^*_{l+4} - \hat{\delta}^*_{l+3} + \frac{1-\alpha_l}{\alpha_{l-1}} \left(\hat{\delta}^*_{l+2} - \frac{1-\alpha_{l-1}}{\alpha_{l-2}} \left(\hat{\delta}^*_{l+1} - (1-\alpha_{l-2})\hat{\delta}^*_l \right) \right) = 0 \tag{7}$$

if knot ξ^*_{l+1} can be removed.

It is seen that removal of knot ξ^*_{l+1} depends on a condition on parameters $\delta^*_l, \ldots, \delta^*_{l+q+1}$. Therefore the leftmost knot that can potentially be checked and removed is ξ_2^*, leaving the first $q+1$ knots unremovable in principle. As a parameter δ_k^* corresponds to knot ξ_{k-q}^*, this means that the first $q+1$ elements of the parameter vector $\delta_1^*, \ldots, \delta_{q+1}^*$ can also not be eliminated. It will be seen that these correspond to the global terms $1, x, x^2, \ldots, x^q$ after transformation to the truncated power series basis (see Section 4), i.e. when all removable knots are eliminated there is still a global polynomial of degree q left, which requires $q+1$ parameters in a model equation. For example, with $q = 2$ it is impossible to arrive at a simple linear component for a model equation, because there will always be an additional quadratic component. Nevertheless, removal of unnecessary knots results in a (considerable) simplification of the model equation.

Checking for knot removal can be done simultaneously by using equations such as (7) for constructing a $(M-q-1) \times M$ matrix P_q that is structured such that the zero elements of $P_q\hat{\delta}^*$ indicate which of the (potentially removable) knots can be eliminated. For example, for $q = 3$ P_q takes the form

$$P_3 = \begin{pmatrix} \ddots & & & & & & \\ & \frac{(1-\alpha_l)(1-\alpha_{l-1})(1-\alpha_{l-2})}{\alpha_{l-1}\alpha_{l-2}} & -\frac{(1-\alpha_l)(1-\alpha_{l-1})}{\alpha_{l-1}\alpha_{l-2}} & \frac{1-\alpha_l}{\alpha_{l-1}} & -1 & \alpha_l & \\ & & & & & & \ddots \end{pmatrix}. \tag{8}$$

3.2 Approximate Knot Removal

For one covariate one might also want to remove a knot when the corresponding element of $P_q\hat{\delta}^*$ is only very close to 0. We therefore extend the knot removal conditions of the type given in (7) and (8) towards approximate knot removal, by introducing a cutoff c for deciding on knot removal. This will allow for a tradeoff between preserving the exact shape of the fitted functions and a simpler model equation based on a small number of basis functions. These minor changes in one variable will hardly have any influence on the other functions in a multivariate model.

The elements of $P_q\hat{\delta}^*$ from a covariance matrix $Cov(\hat{\delta}^*)$ for the initial parameter estimates and the diagonal elements of $P_qCov(\hat{\delta}^*)P_q^T$, combined with a cutoff c, can be used to decide on knot removal. Specifically, a knot $k+1$ and the corresponding parameter $\hat{\delta}_{k+q+1}^*$ will be removed if

$$\frac{[P_q\hat{\delta}^*]_k}{[P_qCov(\hat{\delta}^*)P_q^T]_{kk}} < c \tag{9}$$

where $[]_k$ and $[]_{kk}$ indicate the kth element and the kth diagonal element respectively.

With exact knot removal, i.e. $c = 0$, knots can be removed in any order and it does not matter whether the removal condition is evaluated based on the initial or on the transformed parameter vectors and covariance matrices. This is because removing one knot does not influence the removal condition for the others. There might be a difference for $c > 0$. We propose to remove knots from left to right, because this is the reading direction for the typical interpretation of fitted functions. For evaluation of the knot removal condition two alternative modes will be investigated: In the *simultaneous mode* the knot removal condition (9) is evaluated for every knot simultaneously using the initial covariance matrix. Knots where the condition holds are removed, moving from left to right. In the *stepwise mode* the knot removal condition is evaluated from left to right and knots are removed immediately when the condition holds. After a knot is removed, the transformed covariance matrix (6) is calculated and together with a newly calculated P_q the next knot to be removed is determined.

3.3 Enforcing Knot Removal at Estimation

The P-spline approach of Eilers and Marx (1996) uses a penalty for estimation that is formed by differences of the parameters. In the same way the knot removal condition matrix P_q could be used to construct a penalized likelihood criterion

$$l_P(\delta) = l(\delta) + \lambda \delta^T P_q^T P_q \delta$$

that enforces knot removal at time of estimation. For a B-spline basis of degree q and equally spaced knots it is seen, e.g. from (8), that this is (up to a constant factor) equivalent to a $q+1$th order difference penalty employed in (3) for the P-spline approach. So this approach for penalized estimation will automatically enforce the knot removal condition when the appropriate differencing order for the penalty is employed. This is also true for the mixed model approach used in our example.

While Eilers and Marx (1996) refer to the approximation of an integrated squared curvature penalty for justifying the difference penalty, knot removal enforcement provides a further, exact justification. Furthermore, Eilers and Marx (2004) show that for a large value of the penalty parameter the difference penalty results in fitting of global polynomials, which corresponds to removal of all knots.

4 From B-spline to Truncated Power Series Basis

Compared to B-splines a model equation using a truncated power series basis is expected to be easier to handle, because it only relies on powers and the truncation operator. Eilers and Marx (2004) show how transformation from a B-spline basis to a truncated power series basis can be performed for equally spaced knots. Based on the results of Welham et al. (2007) we extend this to the case of unequally spaced knots, typical after knot removal.

Let B be the $n \times s + q$ matrix that in its columns contains the B-spline basis expansions for all observations. For the corresponding truncated power series basis let F be a $n \times (s-1)$ matrix, that in the columns contains the truncated power functions used by this basis and X a $n \times (q+1)$ matrix that contains the powers $0, \ldots, q$. The truncated power functions in F are only a subset of the truncated power functions needed for calculation of the B-spline basis functions in B (according to (2)). It has to be padded by q columns with additional truncated power functions on the left and by q columns on the right, resulting in the matrix $n \times (s + 2q - 1)$ matrix $\breve{F}$. For performing the transformation $B = \breve{F}\breve{D}_q^T$ the $(s + q) \times (s + 2q + 1)$ matrix $\breve{D}_q$ is needed. Similar to Welham et al. (2007), define the diagonal scaling matrix

$$S_q = diag\left\{\frac{1}{\xi_{l+q} - \xi_l}; l = -q + 1, \ldots, s + 1\right\}$$

and in addition the $(u - 1) \times u$ first order differencing matrices

$$D_u = \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}.$$

Then

$$\breve{D}_q = \begin{cases} D_{s+2}S_1D_{s+3} & \text{for } q = 1 \\ (-1)^{q-1}D_{s+q+1}S_q\breve{D}_{q-1} & \text{for } q > 1 \end{cases}.$$

Let δ be the parameter vector of the B-spline representation and β and b be the parameter vectors of the truncated power series representation, i.e. $B\delta = X\beta + Fb$. Let S be a $(s+4q) \times (s-1)$ matrix formed by a $s-1$ identity matrix bordered by $p + 1$ rows of zeros at the top and at the bottom. Then $F = \breve{F}S$. Let in addition D be given by $D = (\breve{D}_qS)^T$. Following Eilers and Marx (2004) transformation of the parameters is performed by

$$b = D\delta \ , \ \ \beta = (X'X)^{-1}X'(B - FD)\delta.$$

Therefore, for transformation from a B-spline basis to a truncated power series basis transformation matrices are available. This means that, given a covariance matrix for the B-spline parameter estimates (possibly after knot removal), covariance matrices for the truncated power series parameters are available.

5 Example

For illustration of the proposed techniques we use data on children's respiratory health. The data are a subset from the study described in Ihorst et al. (2004), more specifically the data for the German children obtained in autumn 1997. We use only the 811 observations with complete data for the 11 variables considered for modeling. The continuous response of interest is "forced vital capacity" (FVC), a lung function parameter related to children's respiratory health where larger values indicate better lung function. There are 6 continuous covariates for which influence will be modeled by a smooth function: "age" (in years), "birth weight" (bweight), "height", "maximal nitrogen dioxide (NO_2) value of last 24h before lung function measurement" (NOH24), "maximal ozone (O_3) value of last 24h before lung function measurement" (O3H24), and "body mass index" (BMI). Of these the main interest is on "O3H24", because short term ozone exposure is known to have an effect on lung function parameters. In addition, there are 5 binary covariates: "sex" (0=male, 1=female), "whistling or wheezy breath" (whibreath), "shortness of breath, laboured breathing" (shobreath), "assured sensitivity against pollen" (pollsens), and "patient lives in a village with high ozone values" (hiozone). There are no large correlations between covariates. We checked that there are no strong (linear) interactions between covariates which could potentially disturb additive fitting of smooth functions.

5.1 Univariate Model

First we consider a univariate model for "O3H24". Estimation is performed via the mixed model approach described in Wood (2004), employing a B-spline basis of degree $q = 2$ together with a third order difference penalty for enforcing the knot removal condition. Alternatively, we could use cubic basis functions, i.e. $q = 3$, but in our experience there is hardly any difference in the fitted functions and it would lead to more difficult model equations even if all knots can be removed.

We apply knot removal to fits with 10, 20, and 40 initial basis functions to explore the impact of the number of basis functions. Furthermore two modes of evaluation of the knot removal condition are investigated: simultaneous evaluation based on the initial covariance matrix and stepwise evaluation based on the transformed parameter vector and covariance matrix after each removal step. Table 1 shows the resulting number of basis functions/parameters after knot removal using various values for the cutoff c. Figure 1 shows the initial fitted functions (solid curves) together with pointwise confidence intervals for 20 basis functions (dotted curves). The other curves indicate the fits after knot removal using simultaneous evaluation of the knot removal condition (left panel) and stepwise evaluation (right panel) based on various cutoffs c.

It is seen from Table 1 that using a cutoff $c \geq 0.5$ the number of parameters remaining after knot removal hardly depends on the initial number of basis functions. This indicates the latter parameter is rather unimportant with respect to the number of remaining knots and can be chosen from a large range. As seen from the left panel of Figure 1 there are nevertheless differences between the initial fits resulting from various numbers of basis functions, that do not even disappear completely when all knots are removed. For 10 initial basis functions the fitted function seems to have somewhat less complex structure, requiring only 3 instead of 4 parameters after knot removal, which corresponds to a global polynomial of degree 2. We will therefore focus on the fit using 20 initial basis functions.

The simultaneous mode of evaluation of the knot removal condition results in an abrupt drop in the resulting number of parameters and the maximum absolute deviation when the cutoff c is changed from 0.25 to 0.5. The reason for this is seen from the dash-dotted curve (20 basis functions) and the long-

Table 1: Number of remaining parameters (red) and maximum absolute deviation (dev) after knot removal using various cutoffs c and simultaneous (simult) vs. stepwise evaluation (step) of the knot removal condition for a varying initial number (init) of basis functions.

mode	init	cutoff c									
		0.1		0.25		0.5		0.75		1	
		red	dev	red	dev	red	dev	red	dev	red	dev
simult	10	3	0	3	0	3	0	3	0	3	0
	20	10	0.001	8	0.001	3	0.050	3	0.050	3	0.050
	40	17	0.003	11	0.005	3	0.059	3	0.059	3	0.059
step	10	3	0	3	0	3	0	3	0	3	0
	20	5	0.001	4	0.003	4	0.012	4	0.028	3	0.05
	40	14	0.001	4	0.004	4	0.014	4	0.030	3	0.059

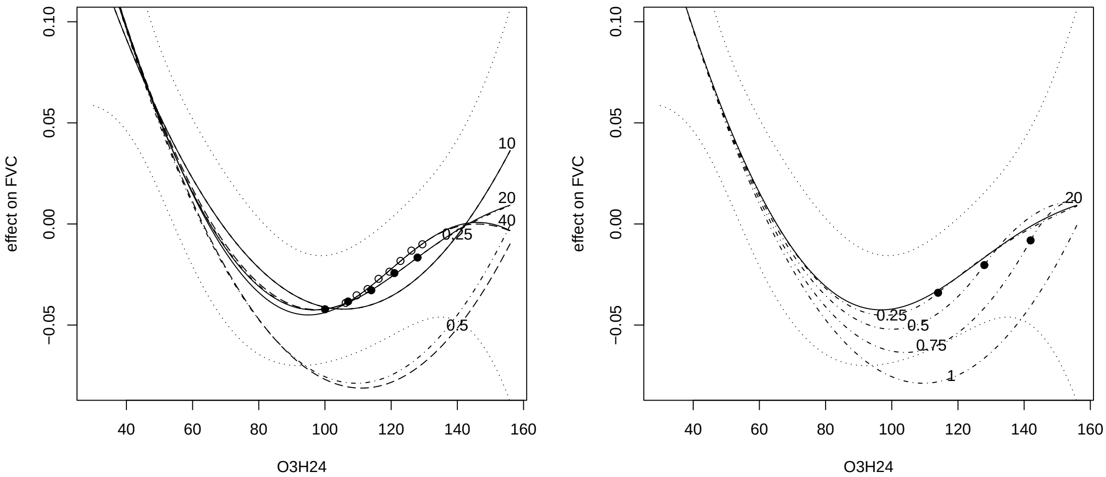

Figure 1: Initial fitted functions (solid curves) and reduced knot fits (dashed/dash-dotted curves) for the univariate lung function model with remaining knots indicated by circles. Pointwise confidence intervals for the original fit using 20 basis functions are indicated by dotted curves. Left panel: simultaneous evaluation of the knot removal condition. For 10 initial basis functions the reduced knot representation coincides with the original fits. For 20 and 40 initial basis functions the reduced knot representations (dash-dotted, filled circles and long-dashed, hollow circles) nearly coincide with the original fit when using $c = 0.25$ and are distinctly different when using $c = 0.5$. Right panel: stepwise evaluation using cutoffs $c \in \{0.25, 0.5, 0.75, 1\}$ and 20 initial basis functions (dash-dotted, filled circles).

dashed curve (40 basis functions) corresponding to $c = 0.25$ in the left panel of Figure 1, which are practically indistinguishable from the original fits, with remaining knots indicated by filled and hollow circles. The latter all have a similar value with respect to the knot removal criterion in (9) and therefore will be eliminated at the same time when the cutoff is increased. This indicates that the simultaneous mode of evaluation critically depends on the exact value of the cutoff, making it rather unstable.

Using the stepwise mode of evaluation of the knot removal condition there is a more gradual decrease of the number of resulting parameters and the maximum absolute deviation. The filled circles on the dashed-dotted curves in the right panel of Figure 1 indicate the one remaining knot within the domain of the covariate for $0.25 \leq c \leq 0.75$. It is seen that this knot moves to the right with increasing value of c, resulting in a worsening of the fits. The reason for this is that, moving from left to right for the evaluation of the knot removal condition, smaller values of c lead to an earlier rejection of the condition. In the present case the structure of the original fitted function is so simple that after that one knot, where the condition does not hold, is retained, all remaining knots can be removed.

As we prefer a gradual decrease of the quality of the fit after knot removal over abrupt changes when the cutoff is increased, we suggest to use the stepwise mode of evaluation of the knot removal condition. In the example the cutoff for the latter can be chosen from a large range and the resulting fits still are within the pointwise confidence intervals. Only for $c = 1$, where no knot remains within the domain of the covariate, the resulting fit is outside these intervals. The resulting global polynomial of degree 2 can no longer adequately represent the structure of the initial fit. Nevertheless, we will use the very cautious value of $c = 0.25$ in the following, which in the example resulted in a fit nearly indistinguishable from the original fit, while reducing the number of parameters from 20 to only 4.

The corresponding truncated power series representation is

$$\begin{aligned} \text{FVC} \quad &= \quad 2.66 - 8.30 \cdot 10^{-3} \cdot \text{O3H24} + 4.24 \cdot 10^{-5} \cdot \text{O3H24}^2 \\ &\quad\; -5.09 \cdot 10^{-5} \cdot (\text{O3H24} - 114)_+^2 + \epsilon \end{aligned}$$

with error term ϵ. Using just one knot instead of 17 knots within the domain of the covariate makes the equation manageable. It can now be used to extract the expected "forced vital capacity" (FVC) by simply plugging in the corresponding value of the covariate "max. O_3 value of last 24h" (O3H24). In addition, the position of the remaining knot indicates in which area there is structure in the data not adequately represented by a global polynomial. For covariate values above 114, i.e. for high O_3 concentrations, the increase starts to level of, which deviates from the u-shape of the global polynomial.

5.2 Multivariate Model

We employ the mixed model approach described in Wood (2004) to fit a multivariate model for "FVC", including 5 binary covariates and 6 smooth functions.

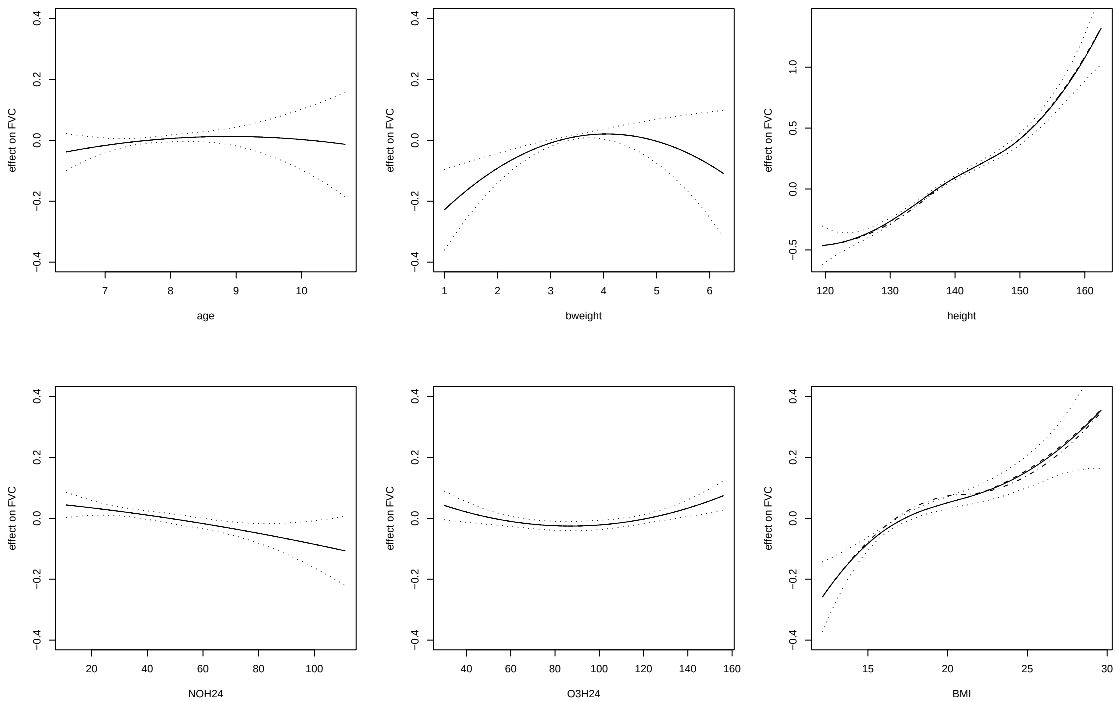

Figure 2: Fitted functions (solid curves) for the multivariate lung function model together with pointwise confidence intervals (dotted curves) and functions resulting from knot reduction with cutoffs $c = 0.25$ (dashed curves) and $c = 1$ (dash-dotted curves). Often the curves are indistinguishable. Note the different scale for the "height" function.

For representation and estimation of each smooth component a B-spline basis of degree 2 with 10 basis functions together with a third order difference penalty is used. Unfortunately, we could not evaluate the use of a larger number of basis functions (e.g. 20 per component), because this resulted in numerical problems. This is a downside of the mixed model approach already encountered in the the simulation study in Binder and Tutz (2006).

Figure 2 shows the resulting fitted smooth functions (solid curves) for the continuous covariates together with approximate pointwise confidence intervals (dotted curves). The covariate "height" seems to have the largest influence on the response (note the different scale), which is plausible because lung volume increases with height.

The functions resulting from knot removal, using cutoff $c = 0.25$, are indicated by dashed curves in Figure 2. Except for a small deviation for large values of covariate "BMI" these are virtually identical to the original fits. At the same time the number of parameters is reduced considerably: For 4 of the 6 continuous covariates ("age", "bweight", "NOH24", and "O3H24") all 7 knots within the domain of the covariate can be removed. That means that only the minimum of 3 basis functions is used for each of these covariates, corresponding to a global polynomial of degree 2. For covariate "height" there are still 7 basis functions and for covariate "BMI" there are 6 basis function. Using a cutoff of $c = 1$ reduces this to 6 and 4 basis functions respectively, but the corresponding fits (dash-dotted curves in Figure 2) already start to differ noticeably from the original fits. Therefore it seems that the use of 7 and 6 basis functions respectively is justified. Taken together, there is a considerable simplification compared to the initial 10 basis functions used for representation of each of the fitted smooth functions.

The corresponding coefficients of the truncated power series representation are given in Table 2. In addition the standard error estimates based on the transformed covariance matrices are given.

The constant terms from all components have been absorbed into the intercept term and therefore no standard error is given for the latter. The approximate standard error estimates are not intended for selection of components, which should be based on p-values for the whole fitted functions. (Mis-)using the standard errors for significance tests would result in stepwise knot selection, which is not the intention of the present approach.

For the truncated power representation resulting from $c = 0.25$ there are still 25 parameters to be considered for the prediction of a new observation. Nevertheless, this presents a considerable simplification over the original fit which would correspond to 60 parameters (with the constant terms being absorbed into the intercept term). Note also that 5 of the 25 parameters belong to binary covariates and therefore were not even subject to reduction. So only a mean number of about 3 parameters is used for representing the smooth influence of a continuous covariate, which is close to the number typically used in unpenalized approaches. Also, a simple linear model would already have 12 parameters and a model that features quadratic components 18 parameters. Taken together, the reduced representation is much improved with respect to transportability, while preserving the shape of the fitted functions and thereby predictive performance.

One indication where interesting local features of the fitted functions are to be found is given by the position of the knots remaining after knot removal. This does not mean that there is no change of the fitted function in other areas as the shape there still corresponds to a polynomial of degree 2, but only that in areas with a larger number of knots global polynomials are no longer sufficient to adequately describe the structure. For example for "height" the knot positions in the representation resulting from using $c = 0.25$ together with the full sample fit indicate that important local structure is present for values between 130 and 146. This corresponds to the increase in slope up to about 136 and the decrease in slope up to 146 seen in Figure 2.

For the representation resulting from using $c = 1$ all parameters except the ones for "height" and "BMI" stay the same. The complexity of the representation for the "BMI" component is reduced considerably. The parameter estimate

Table 2: Coefficients and their standard errors for the truncated power series representation of the reduced knot versions of the multivariate lung function model fitted to the full data of size $n = 811$ (using $c = 0.25$ and $c = 1$) and to a sub-sample of size $n = 405$ (using $c = 0.25$).

	$n = 811$, $c = 0.25$		$n = 811$, $c = 1$		$n = 405$, $c = 0.25$	
	coef	SE	coef	SE	coef	SE
intercept	$1.84 \cdot 10^{1}$	-	$1.65 \cdot 10^{1}$	-	1.66	-
sex	$-1.83 \cdot 10^{-1}$	$1.50 \cdot 10^{-2}$	$-1.83 \cdot 10^{-1}$	$1.50 \cdot 10^{-2}$	$-2.01 \cdot 10^{-1}$	$2.17 \cdot 10^{-2}$
whibreath	$3.17 \cdot 10^{-2}$	$3.85 \cdot 10^{-2}$	$3.17 \cdot 10^{-2}$	$3.85 \cdot 10^{-2}$	$4.07 \cdot 10^{-2}$	$5.36 \cdot 10^{-2}$
shobreath	$1.03 \cdot 10^{-1}$	$4.01 \cdot 10^{-2}$	$1.03 \cdot 10^{-1}$	$4.01 \cdot 10^{-2}$	$1.32 \cdot 10^{-1}$	$6.13 \cdot 10^{-2}$
pollsens	$-4.40 \cdot 10^{-2}$	$1.78 \cdot 10^{-2}$	$-4.40 \cdot 10^{-2}$	$1.78 \cdot 10^{-2}$	$-4.63 \cdot 10^{-2}$	$2.55 \cdot 10^{-2}$
hiozone	$-5.43 \cdot 10^{-2}$	$2.09 \cdot 10^{-2}$	$-5.43 \cdot 10^{-2}$	$2.09 \cdot 10^{-2}$	$-6.33 \cdot 10^{-2}$	$3.07 \cdot 10^{-2}$
age	$1.46 \cdot 10^{-1}$	$1.92 \cdot 10^{-1}$	$1.46 \cdot 10^{-1}$	$1.92 \cdot 10^{-1}$	$2.28 \cdot 10^{-1}$	$2.80 \cdot 10^{-1}$
age^2	$-8.21 \cdot 10^{-3}$	$1.20 \cdot 10^{-2}$	$-8.21 \cdot 10^{-3}$	$1.20 \cdot 10^{-2}$	$-1.32 \cdot 10^{-2}$	$1.74 \cdot 10^{-2}$
bweight	$2.16 \cdot 10^{-1}$	$7.94 \cdot 10^{-2}$	$2.16 \cdot 10^{-1}$	$7.94 \cdot 10^{-2}$	$1.76 \cdot 10^{-1}$	$1.40 \cdot 10^{-1}$
bweight^2	$-2.67 \cdot 10^{-2}$	$1.18 \cdot 10^{-2}$	$-2.67 \cdot 10^{-2}$	$1.18 \cdot 10^{-2}$	$-2.27 \cdot 10^{-2}$	$2.15 \cdot 10^{-2}$
height	$-3.27 \cdot 10^{-1}$	$3.31 \cdot 10^{-1}$	$-2.92 \cdot 10^{-1}$	$2.57 \cdot 10^{-1}$	$-6.09 \cdot 10^{-2}$	$4.89 \cdot 10^{-2}$
height^2	$1.38 \cdot 10^{-3}$	$1.30 \cdot 10^{-3}$	$1.24 \cdot 10^{-3}$	$9.85 \cdot 10^{-4}$	$3.38 \cdot 10^{-4}$	$1.77 \cdot 10^{-4}$
(height-130)$_+^2$	$-8.72 \cdot 10^{-4}$	$1.99 \cdot 10^{-3}$	-	-	-	-
(height-136)$_+^2$	$-1.62 \cdot 10^{-3}$	$1.23 \cdot 10^{-3}$	$-2.78 \cdot 10^{-3}$	$2.03 \cdot 10^{-3}$	-	-
(height-141)$_+^2$	$1.25 \cdot 10^{-3}$	$1.11 \cdot 10^{-3}$	$1.69 \cdot 10^{-3}$	$1.46 \cdot 10^{-3}$	-	-
(height-146)$_+^2$	$2.06 \cdot 10^{-3}$	$1.21 \cdot 10^{-3}$	$2.06 \cdot 10^{-3}$	$1.21 \cdot 10^{-3}$	-	-
NOH24	$-8.84 \cdot 10^{-4}$	$1.57 \cdot 10^{-3}$	$-8.84 \cdot 10^{-4}$	$1.57 \cdot 10^{-3}$	$-8.28 \cdot 10^{-4}$	$2.18 \cdot 10^{-3}$
NOH24^2	$-5.09 \cdot 10^{-6}$	$1.53 \cdot 10^{-5}$	$-5.09 \cdot 10^{-6}$	$1.53 \cdot 10^{-5}$	$-4.31 \cdot 10^{-6}$	$2.08 \cdot 10^{-5}$
O3H24	$-3.63 \cdot 10^{-3}$	$1.25 \cdot 10^{-3}$	$-3.63 \cdot 10^{-3}$	$1.25 \cdot 10^{-3}$	$-5.16 \cdot 10^{-3}$	$1.81 \cdot 10^{-3}$
O3H24^2	$2.09 \cdot 10^{-5}$	$6.64 \cdot 10^{-6}$	$2.09 \cdot 10^{-5}$	$6.64 \cdot 10^{-6}$	$2.86 \cdot 10^{-5}$	$9.46 \cdot 10^{-6}$
BMI	$1.98 \cdot 10^{-1}$	$1.15 \cdot 10^{-1}$	$1.80 \cdot 10^{-1}$	$8.83 \cdot 10^{-2}$	$1.64 \cdot 10^{-1}$	$1.10 \cdot 10^{-1}$
BMI^2	$-5.02 \cdot 10^{-3}$	$3.57 \cdot 10^{-3}$	$-4.27 \cdot 10^{-3}$	$2.45 \cdot 10^{-3}$	$-3.87 \cdot 10^{-3}$	$3.19 \cdot 10^{-3}$
(BMI-16.5)$_+^2$	$1.62 \cdot 10^{-3}$	$4.43 \cdot 10^{-3}$	-	-	-	-
(BMI-18.7)$_+^2$	$2.68 \cdot 10^{-3}$	$2.92 \cdot 10^{-3}$	-	-	$2.31 \cdot 10^{-3}$	$5.57 \cdot 10^{-3}$
(BMI-20.9)$_+^2$	$2.99 \cdot 10^{-3}$	$2.76 \cdot 10^{-3}$	$7.70 \cdot 10^{-3}$	$4.16 \cdot 10^{-3}$	$2.53 \cdot 10^{-3}$	$3.62 \cdot 10^{-3}$
(BMI-23.1)$_+^2$	-	-	-	-	$1.52 \cdot 10^{-3}$	$3.22 \cdot 10^{-3}$

for one remaining knot within the domain of the covariate is larger than the corresponding estimate in the representation for $c = 0.25$. So it seems that this one knot takes over the role of three knots from the latter representation. This results in a distinct change of the fitted function. For the "height" component all but one knot are retained compared to the representation for $c = 0.25$. This is probably due to the narrow confidence intervals and the corresponding covariance matrix which result in a more strict knot removal criterion. So it seems that the local structure of the "height" function is well supported by the data.

To explore the influence of sample size on the fitted model and on the reduced knot representation we fit the multivariate model to a random subsample of the data of size $n = 405$. Figure 3 shows the fitted functions (dashed curves) after knot reduction (using $c = 0.25$) and Table 2 shows the corresponding coefficients. It is seen that, compared to the fit from the full sample (solid curves), the local structure for the "height" function has disappeared. The reduced representation does not feature any knot within the domain of the covariate. This might cast doubt on the validity of the local features found for the "height" function in the full sample. For the "BMI" component the fits from the full sample and from the subsample look very similar and for the latter still three knots are retained within the domain of the covariate (featuring slightly different positions). This might suggest validity of the local features of the "BMI" function. For all other covariates the fitted functions and parameter estimates are very similar, indicating robustness of the fitted model and of the reduced knot representation.

6 Discussion

Modern techniques for fitting generalized additive models can deal with a large number of covariates. For example, in the present paper we constructed a model for children's lung function featuring 6 continuous covariates with potentially non-linear influence in addition to 5 binary covariates. The downside is that, due to the large number of basis functions, the results can only be plotted, but the corresponding model equation is much to complicated to be transported. In our example it would have contained 60 parameters.

One prominent way for representing smooth function is the B-spline basis. With this basis, or a basis that can be transformed to the B-spline basis, knots can often be removed without changing the shape of the function. As this knot removal is implemented without refitting as a linear transform of the (estimated) parameters, also the covariance matrix of the parameter estimates can be transformed. Thereby simplified model equations including standard error estimates are obtained. We extended this to approximate knot removal where a certain margin for the knot removal condition is allowed. It was demonstrated that this can result in a considerably simplified model equation while closely preserving shape. In our example we obtained a representation with only 25 parameters compared to the original 60 parameters, where several components could be reduced to a global polynomial form. In this simplified form the model can then easily used by other researchers to plug in covariate values of their

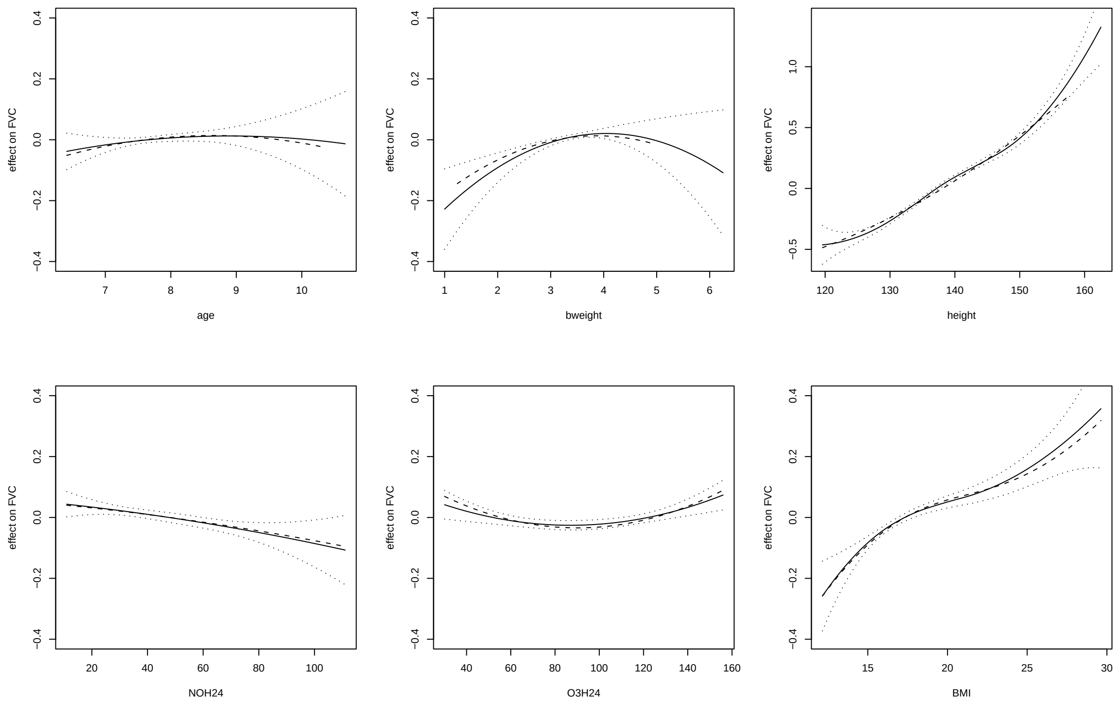

Figure 3: Fitted functions resulting from knot removal with cutoff $c = 0.25$ for the multivariate lung function model fitted to the full data (solid curves) and to a subsample of size $n = 405$ (dashed curves), together with pointwise confidence intervals for the model from the full data (dotted curves).

observational units to derive predictions from their own data or by doctors to calculate predictions for patients. Such an application of a fitted model on new data is essential for validation and might help to finally derive a robustified and generally applicable model.

Decreasing the sample size, at least for the present example, resulted in a model with less local structure. While this might be seen as a loss of information with respect to the underlying structure, this also lead to a simpler, maybe more robust, model fit with a smaller number of parameters in the reduced knot representation. In any case, the knot removal technique allowed to explore the effects of sample size on local features more closely.

The proposed knot removal approach could potentially also be useful for analysing stability of fitted generalized additive models. When the fitting procedure under investigation is repeatedly applied to bootstrap samples, knot reduction in every bootstrap samples could indicate which local features are consistently retained and which ones may be just artefacts.

One use of the knot removal technique that has not been illustrated in this paper is that of comparison of model fits. There are various approaches for fitting generalized additive models and so far the results from different approaches are often only compared by inspecting the plots of the fitted functions. Given that each approach uses the B-spline basis or a basis that can be transformed to it (such as the truncated power series basis) reduced knot representations for the fits from the single approaches can be obtained. Given that the number of knots can be reduced considerably, these representations can then be more easily compared between the approaches. Recently, Govindarajulu et al. (2007) presented an alternative approach where the (weighted) area between fitted functions is used for comparison.

One potential further extension is with respect to the set of global functions: When all potentially removable knots are removed in the present approach, global polynomials are obtained. This is only a very limited class of global functions. When there is only little local structure, the fractional polynomial approach (Royston and Altman, 1994; Sauerbrei and Royston, 1999) is more flexible, resulting in a fit that is at least as good as a global polynomial. It also has the advantage of selecting a simple linear function if a quadratic term does not improve the fit. We are currently investigating new approaches that allow for a larger set of global functions in combination with a (small) number of local features obtained by knot removal.

Finally, the approach is not limited to continuous response models. Whenever a generalized additive model can be fitted (e.g. for a binary or a Poisson response) using a B-spline basis or a basis that can be transformed to the B-spline basis, knot removal can be applied afterwards. This is also the case for even more complex model classes such as generalized additive mixed models. For example for the lung function data used in this paper there are additional repeated measures available. Therefore the next step in model building could be to construct an appropriate mixed model. The resulting representation could then again be simplified by knot removal. This shows that the proposed approach is a general tool that is applicable in a wide variety of settings.

Acknowledgements

We gratefully acknowledge support from Deutsche Forschungsgemeinschaft (SA 580/4-1). We also thank Gabi Ihorst for providing the ozone study data.

References

Binder, H., Tutz, G., 2006. Fitting generalized additive models: A comparison of methods. FDM-Preprint 93, University of Freiburg.

Boehm, W., 1980. Inserting new knots into B-spline curves. Computer-Aided Design 12 (4), 199–201.

Breiman, L., 1993. Fitting additive models to regression data. Computational Statistics & Data Analysis 15, 13–46.

de Boor, C., 2001. A Practical Guide to Splines (Revised Edition). Springer, New York.

Eilers, P. H. C., Marx, B. D., 1996. Flexible smoothing with B-splines and penalties. Statistical Science 11 (2), 89–121.

Eilers, P. H. C., Marx, B. D., 2004. Splines, knots, and penalties, manuscript, submitted.

Friedman, J. H., 1991. Multivariate adaptive regression splines. The Annals of Statistics 19 (1), 1–67.

Friedman, J. H., Silverman, B. W., 1989. Flexible parsimonious smoothing and additive modeling. Technometrics 31 (1), 3–21.

Gervini, D., 2006. Free-knot spline smoothing for functional data. Journal of the Royal Statistical Society B 68 (4), 671–688.

Govindarajulu, U. S., Spiegelman, D., Thurston, S. W., Ganguli, B., Eisen, E. A., 2007. Comparing smoothing techniques in Cox models for exposure-response relationships. Statistics in Medicine 26 (20), 3735–3752.

Green, P. J., Silverman, B. W., 1994. Nonparametric Regression and Generalized Linear Models. Chapman & Hall, London.

Hastie, T. J., Tibshirani, R. J., 1990. Generalized Additive Models. Chapman & Hall, London.

Ihorst, G., Frischer, T., Horak, F., Schumacher, M., Kopp, M., Forster, J., Mattes, J., Kuehr, J., 2004. Long- and medium-term ozone effects on lung growth including a broad spectrum of exposure. European Respiratory Journal 23, 292–299.

Kooperberg, C., Bose, S., Stone, C. J., 1997. Polychotomous regression. Journal of the American Statistical Association 92 (437), 117–127.

Mao, W., Zhao, L. H., 2003. Free-knot polynomial splines with confidence intervals. Journal of the Royal Statistical Society B 65 (4), 901–919.

Marx, B. D., Eilers, P. H. C., 1998. Direct generalized additive modelling with penalized likelihood. Computational Statistics and Data Analysis 28, 193–209.

Miyata, S., Shen, X., 2003. Adaptive free-knot splines. Journal of Computational & Graphical Statistics 12 (1), 197–213.

Molinari, N., Durand, J.-F., Sabatier, R., 2004. Bounded optimal knots for rgression splines. Computational Statistics & Data Analysis 45, 159–178.

Royston, P., Altman, D. G., 1994. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. Applied Statistics 43 (3), 429–467.

Ruppert, D., 2002. Selecting the number of knots for penalized splines. Journal of Computational and Graphical Statistics 11, 735–757.

Ruppert, D., Wand, M. P., Carroll, R. J., 2003. Semiparametric Regression. Cambridge University Press.

Sauerbrei, W., Royston, P., 1999. Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. Journal of The Royal Statistical Society A 162 (1), 71–94.

Speed, T., 1991. Comment on "That BLUP is a good thing: The estimation of random effects" by G. K. Robinson. Statistical Science 6 (1), 42–44.

Stone, C. J., Hansen, M. H., Kooperberg, C., Truong, Y. K., 1997. Polynomial splines and their tensor products in extended linear modeling: 1994 Wald memorial lecture. Annals of Statistics 25 (4), 1371–1470.

Tiller, W., 1992. Knot-removal algorithms for NURBS curves and surfaces. Computer-Aided Design 24 (8), 445–453.

Welham, S. J., Cullis, B. R., Kenward, M. G., Thompson, R., 2007. A comparison of mixed model splines for curve fitting. Australian & New Zealand Journal of Statistics 49 (1), 1–23.

Wood, S. N., 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association 99 (467), 673–686.

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

Wyatt, J. C., Altman, D. G., 1995. Commentary: Prognostic models: clinically useful of quickly forgotten. British Medical Journal 311, 1539–1541.

Zhou, S., Shen, X., 2001. Spatially adaptive regression splines and accurate knot selection schemes. Journal of the American Statistical Association 96 (453), 247–259.