BIOMETRICS 57, 539–545
June 2001

Simple Incorporation of Interactions into Additive Models

Brent A. Coull,[1,*] David Ruppert,[2] and M. P. Wand[1]

[1]Department of Biostatistics, Harvard School of Public Health,
655 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.
[2]School of Operations Research and Industrial Engineering, Cornell University,
Ithaca, New York 14853, U.S.A.
*email: bcoull@hsph.harvard.edu

SUMMARY. Often, the functional form of covariate effects in an additive model varies across groups defined by levels of a categorical variable. This structure represents a factor-by-curve interaction. This article presents penalized spline models that incorporate factor-by-curve interactions into additive models. A mixed model formulation for penalized splines allows for straightforward model fitting and smoothing parameter selection. We illustrate the proposed model by applying it to pollen ragweed data in which seasonal trends vary by year.

KEY WORDS: Generalized additive model; Generalized linear mixed model; Penalized spline; Pollen forecasting; Varying-coefficient model.

1. Introduction

An additive model (Hastie and Tibshirani, 1990) expresses the mean of a response variable Y as a sum of low-dimensional smooth functions of covariates. In their simplest form, these models express covariate effects as univariate functions; i.e., the functional form for the effect of a covariate on the response does not depend on the values of other covariates. In practice, however, the functional form of a covariate effect often varies according to the values taken by one or more of the remaining covariates, making this additivity assumption untenable.

Existing work on curve interactions in additive models has focused mainly on interactions involving continuous covariates. A simple case of this type of interaction is a bivariate function of two covariates (cf., Hastie and Tibshirani, 1990, pp. 264–278; Hobert, Altman, and Schofield, 1997). More generally, Wahba (1986, 1988) and Chen (1987) discussed ANOVA-like methods in which the response is successively modeled as linear combinations of univariate functions, linear combinations of univariate and bivariate functions, linear functions of univariate, bivariate, and trivariate functions, and so on. These authors called the resulting estimates main effect splines, two-factor interaction splines, three-factor interaction splines, and so on. Chen (1993) used smoothing splines to fit models containing terms of arbitrary order k.

In many cases, one might expect the form of a functional relationship between the mean of Y and one or more covariates to vary among subsets of observations defined by levels of a categorical predictor Z. This structure represents a factor-by-curve interaction. Figure 1 shows a simple example of a factor-by-curve interaction in which the mean of the response Y can be expressed as a smooth function of a single covariate X, with the form of this function depending on group membership denoted by the binary variable Z. (See Hastie and Tibshirani (1990, pp. 265–266) for additional examples of models containing factor-by-curve interactions.)

Algorithms for fitting factor-by-curve interactions have received relatively little attention. To our knowledge, none of the existing commercial smoothing software can fit such models without significant additional programming effort. For instance, Chambers and Hastie (1993, pp. 269–270) noted that the S-PLUS function for fitting generalized additive models, gam, does not currently support factor-by-curve interactions. In order to fit such a model in S-PLUS, one must program a backfitting algorithm in which, for each backfit iteration corresponding to a curve-by-interaction term, one splits the data according to the levels of z and fits a smooth function to each subset. (See Coull, Catalano, and Godleski (2000) for an example of this approach.) When the number of factor-by-curve interaction terms or the number of factor levels becomes large, this backfitting approach becomes cumbersome. Moreover, traditional methods for selecting the smoothing parameters, such as generalized cross validation, are awkward because of the need to minimize a particular criterion over the multivariate smoothing parameter space.

We propose the use of penalized spline models (Eilers and Marx, 1996) as a simple way of incorporating factor-by-curve interactions into generalized additive models. Since fitting penalized splines does not involve backfitting, the methods are trivial to program. In particular, a mixed model formulation of penalized spline models allows one to fit the models using existing mixed model software, such as the SAS procedure PROC MIXED (Littell et al., 1996) or S-PLUS function lme (Venables and Ripley, 1994). Moreover, smoothing parameter selection is a by-product of model fitting.

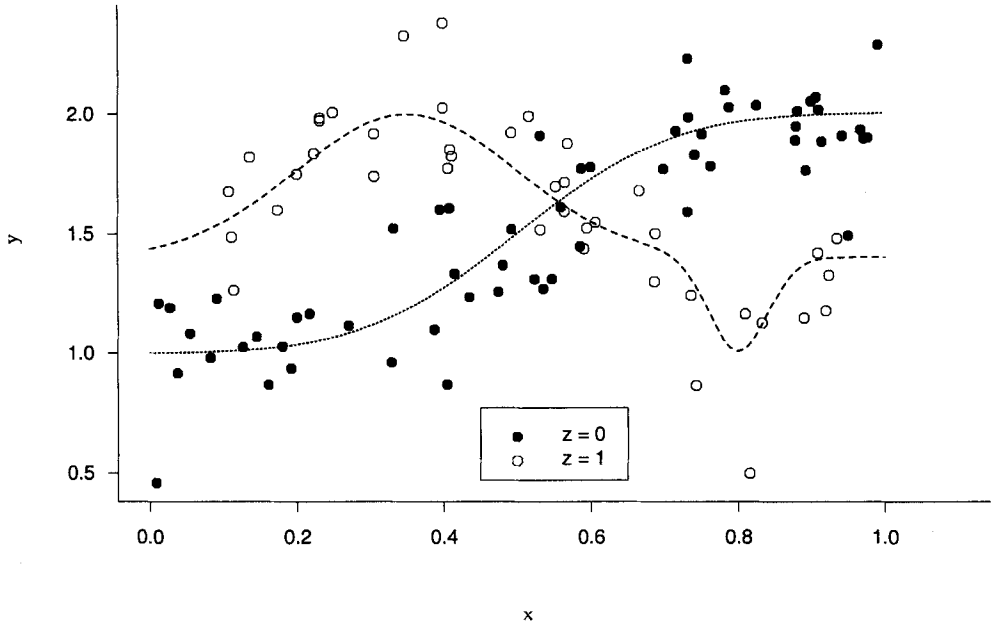

Figure 1. Simple example of a factor-by-curve interaction.

We use the proposed methods to reanalyze data on daily ragweed pollen counts. Since avoidance plays a large role in the treatment of pollen-related allergies, a major objective in aerobiology is the development of accurate forecasting models for daily pollen levels (Stark et al., 1997). Figure 2 shows data on daily ragweed pollen counts from four consecutive pollen seasons in Michigan. Stark et al. (1997) and Brumback et al. (2000) fit generalized linear and generalized additive models, respectively, to these data to investigate the predictive power of meteorological variables on pollen level. Because each year's pollen season starts at a different time and will progress at a different rate, the seasonal trend is thought to be different for each year. This year-to-year heterogeneity led both sets of authors to fit models to each year's data separately. Section 7 presents a single analysis based on data from all four pollen seasons.

Sections 2 and 3 present penalized spline models for factor-by-curve interactions. Section 4 discusses standard error and degrees of freedom calculations, and Section 5 discusses smoothing parameter selection. Section 6 incorporates the methods into generalized additive models. Section 7 applies the methods to the pollen data and is followed by discussion in Section 8.

2. Penalized Splines and Interactions

For simplicity, we first explain factor-by-curve interactions for a single continuous predictor and a single categorical factor. Consider the set of triples (x_i, y_i, z_i), $1 \le i \le n$, where the x_i and y_i represent continuous predictor and response recordings, respectively, and $z_i \in \{1, \ldots, L\}$ represents a coded factor. The type of model that we wish to fit is

$$y_i = f_{z_i}(x_i) + \varepsilon_i, \qquad 1 \le i \le n \tag{1}$$

where $f_1, \ldots, f_L$ are L different functions depending on the value of z_i and ε_i i.i.d. $N(0, \sigma_\varepsilon^2)$. Figure 1 shows an example of data generated from model (1) with $L = 2$ and the true curves f_1 and f_2 denoted by dashed lines.

Let $\kappa_1, \ldots, \kappa_K$ be a set of distinct numbers, or knots, inside the range of the x_i's and let $x_+ = \max(0, x)$. The knots are usually taken to be relatively dense among the observations in an attempt to capture the curvature in f_ℓ, $\ell = 1, \ldots, L$. A reasonable allocation rule is one knot for every four to five observations, up to a maximum of about 40 knots. Ruppert and Carroll (2000) described an algorithm for choosing the number of knots and demonstrated its effectiveness through simulation. Define

$$z_{i\ell} = \begin{cases} 1 & \text{if } z_i = \ell, \\ 0 & \text{otherwise,} \end{cases}$$

for $\ell = 1, \ldots, L$. A linear penalized spline model (Eilers and Marx, 1996) for (1) is

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} b_k (x_i - \kappa_k)_+ + \sum_{\ell=2}^{L} z_{i\ell} \left(\gamma_{0\ell} + \gamma_{1\ell} x_i\right) + \sum_{\ell=1}^{L} z_{i\ell} \left\{ \sum_{k=1}^{K} c_k^\ell (x_i - \kappa_k)_+ \right\} + \varepsilon_i, \tag{2}$$

subject to the constraints

$$\sum_{k=1}^{K} b_k^2 < B \quad \text{and} \quad \sum_{k=1}^{K} \left(c_k^\ell\right)^2 < C_\ell, \qquad \ell = 1, \ldots, L, \tag{3}$$

for some constants B and C_ℓ, $\ell = 1, \ldots, L$. In model (2), $(\gamma_{0\ell} + \gamma_{1\ell} x_i)$ models the linear deviation between f_1 and f_ℓ, $\ell = 2, \ldots, L$, whereas $\Sigma_{k=1}^K c_k^\ell (x_i - \kappa_k)_+$ represents deviations from the overall smooth term $\Sigma_{k=1}^K b_k (x_i - \kappa_k)_+$. The penalty (3) induces smoothness in the effect of x on y so that the exact

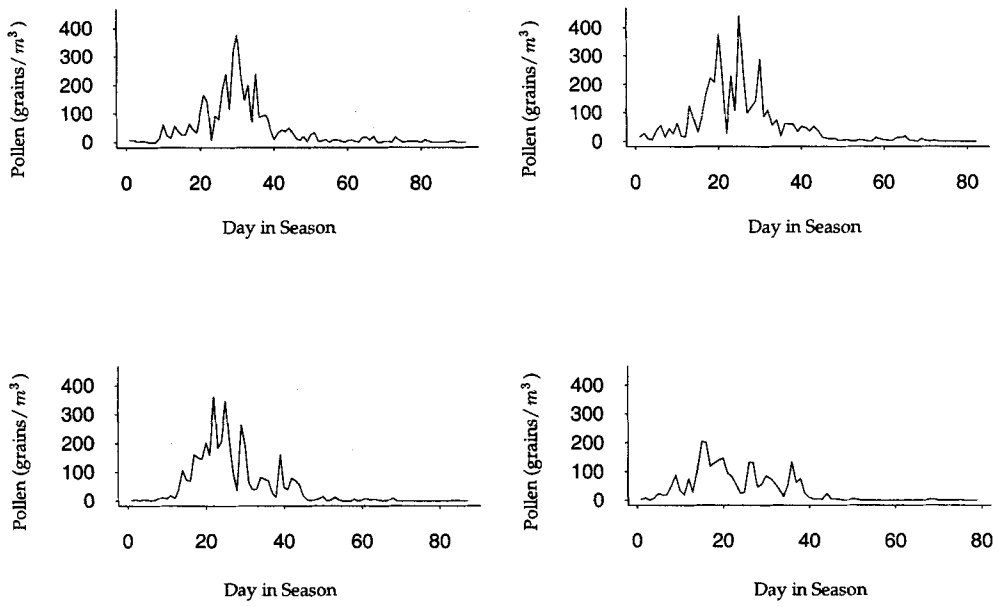

Figure 2. Total daily ragweed pollen counts for 1991–1994 pollen seasons in Kalamazoo, Michigan.

number of knots is not a major concern provided enough have been specified.

Brumback, Ruppert, and Wand (1999) pointed out that, for given values of B and C_ℓ, $\ell = 1, \ldots, L$, model (2) subject to constraints (3) yields fitted values equivalent to those produced by the model

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} b_k (x_i - \kappa_k)_+ + \sum_{\ell=2}^{L} z_{i\ell} (\gamma_{0\ell} + \gamma_{1\ell} x_i) + \sum_{\ell=1}^{L} z_{i\ell} \left\{ \sum_{k=1}^{K} c_k^\ell (x_i - \kappa_k)_+ \right\} + \varepsilon_i, \tag{4}$$

where b_k i.i.d. $\mathrm{N}(0, \sigma_b^2)$, and c_k^ℓ i.i.d. $\mathrm{N}(0, \sigma_{c\ell}^2)$, $\ell = 1, \ldots, L$, for appropriate values of σ_b and $\sigma_{c\ell}$. Henceforth, we use this mixed model formulation of penalized spline models. We can write model (4) in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \tag{5}$$

where

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \gamma_{02}, \ldots, \gamma_{0L}, \gamma_{12}, \ldots, \gamma_{1L})^{\mathrm{T}},$$
$$\mathbf{u} = (b_1, \ldots, b_K, c_1^1, c_2^1, \ldots, c_K^1, c_1^2, \ldots, c_K^L)^{\mathrm{T}},$$
$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & z_{12} & \cdots & z_{1L} & z_{12}x_1 & z_{13}x_1 & \cdots & z_{1L}x_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & z_{n2} & \cdots & z_{nL} & z_{n2}x_n & z_{n3}x_n & \cdots & z_{nL}x_n \end{bmatrix},$$
$$\mathbf{Z} = \begin{bmatrix} (x_1 - \kappa_1)_+ & \cdots & (x_1 - \kappa_K)_+ & z_{11}(x_1 - \kappa_1)_+ & \cdots & z_{11}(x_1 - \kappa_K)_+ & \cdots & z_{1L}(x_1 - \kappa_K)_+ \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ (x_n - \kappa_1)_+ & \cdots & (x_n - \kappa_K)_+ & z_{n1}(x_n - \kappa_1)_+ & \cdots & z_{n1}(x_n - \kappa_K)_+ & \cdots & z_{nL}(x_n - \kappa_K)_+ \end{bmatrix},$$

and

$$\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I} \end{bmatrix}\right),$$

with $\mathbf{G} = \mathrm{diag}(\sigma_b^2 \mathbf{1}_K, \sigma_{c1}^2 \mathbf{1}_K, \ldots, \sigma_{cL}^2 \mathbf{1}_K)$. Here, $\mathbf{1}_K$ is the $K \times 1$ vector of ones. Thus, penalized spline model (4) falls within the linear mixed model framework, and we can rely on the well-developed body of methodology for this broad class of models. In particular, the best linear unbiased predictor (BLUP) of $\mathbf{y}$ (Robinson, 1991) is

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}},$$

where

$$\hat{\boldsymbol{\beta}} = \left\{ \mathbf{X}^{\mathrm{T}} \left(\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \sigma_\varepsilon^2 \mathbf{I} \right)^{-1} \mathbf{X} \right\}^{-1} \mathbf{X}^{\mathrm{T}} \left(\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \sigma_\varepsilon^2 \mathbf{I} \right)^{-1} \mathbf{y} \tag{6}$$

and

$$\hat{\mathbf{u}} = \sigma_\varepsilon^2 \left(\sigma_\varepsilon^2 \mathbf{Z}^{\mathrm{T}}\mathbf{Z} + \mathbf{G}^{-1} \right)^{-1} \mathbf{Z}^{\mathrm{T}} \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right). \tag{7}$$

Extension to models with truncated polynomials $(x_i-\kappa_k)_+^p$ for $p>1$ is straightforward. Specifically, the pth order penalized spline model for (1) is

$$y_i = \beta_0+\beta_1 x_i+\beta_2 x_i^2+\cdots+\beta_p x_i^p + \sum_{k=1}^{K} b_k(x_i-\kappa_k)_+^p + \sum_{\ell=2}^{L} z_{i\ell}\left(\gamma_{0\ell}+\gamma_{1\ell}x_i+\cdots+\gamma_{p\ell}x_i^p\right) + \sum_{\ell=1}^{L} z_{i\ell}\left\{\sum_{k=1}^{K} c_k^{\ell}(x_i-\kappa_k)_+^p\right\}+\varepsilon_i,$$

where again b_k i.i.d. $N(0,\sigma_b^2)$ and c_k^{ℓ} i.i.d. $N(0,\sigma_{c\ell}^2)$, $\ell=1,\ldots,L$.

3. Additive Models with Interactions

Model (4) specifies a different smooth function $f(x_i)$ for each subset of observations defined by the levels of z. Thus, one can effectively fit this model (apart from the assumption of homoskedastic errors across factor levels) by fitting a nonparametric regression model to each subset separately. For a multiple regression model in which some terms do not interact with z, however, the penalized spline approach holds a substantial advantage over the data subsetting approach since the latter must be nested within a backfitting algorithm.

To keep notation simple, we now consider a semiparametric model with a single parametric term, a single nonparametric term, and a single factor-by-curve interaction. Extension to models with more than one term of each type is straightforward. Consider now the multiple regression setting with response y_i, general predictor x_i, continuous predictors (s_i,t_i), and categorical predictor z_i, $i=1,\ldots,n$. A semiparametric model for y_i that allows the functional form of the effect of t_i on y_i to vary according to the level of z_i is

$$y_i=\alpha_0+\alpha_1 x_i+g(s_i)+f_{z_i}(t_i)+\varepsilon_i, \qquad 1\le i\le n. \quad (8)$$

To generalize the arguments in Section 2, let $\kappa_1^s,\ldots,\kappa_{K_s}^s$ and $\kappa_1^t,\ldots,\kappa_{K_t}^t$ be the K_s and K_t knots corresponding to s_i and t_i, respectively. In addition, let $z_{i\ell}$, $i=1,\ldots,n$, $\ell=1,\ldots,L$, be defined as in Section 2. A linear penalized spline model for (8) is

$$y_i=\alpha_0+\alpha_1 x_i+\beta_1^s s_i+\sum_{k=1}^{K_s} b_k^s(s_i-\kappa_k^s)_+ +\beta_1^t t_i + \sum_{k=1}^{K_t} b_k^t(t_i-\kappa_k^t)_+ + \sum_{\ell=2}^{L} z_{i\ell}(\gamma_{0\ell}+\gamma_{1\ell}t_i) + \sum_{\ell=1}^{L} z_{i\ell}\left\{\sum_{k=1}^{K_t} c_k^{\ell}(t_i-\kappa_k^t)_+\right\}+\varepsilon_i, \quad (9)$$

where b_k^s i.i.d. $N(0,\sigma_{bs}^2)$, b_k^t i.i.d $N(0,\sigma_{bt}^2)$, and c_k^{ℓ} i.i.d. $N(0,\sigma_{c\ell}^2)$. Model (9) also falls within the mixed model framework (5), making estimation and inference no more difficult than that for the single covariate model (4).

4. Standard Error and Degrees of Freedom Calculations

Variability bands in function estimation are usually obtained by adding and subtracting twice the estimated standard error of the estimated function (e.g., Bowman and Azzalini, 1997, pp. 75–76). Bias aside, they can be interpreted as approximate pointwise confidence intervals (Hastie and Tibshirani, 1990). They are also useful for detection of leverage and display of inherent variability. For additive models in the linear mixed model framework, the standard errors are easily derived using standard multivariate statistical manipulations after obtaining an estimate of $\mathrm{cov}([\hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\mathbf{u}}^{\mathrm{T}}]^{\mathrm{T}}|\mathbf{u})$. However, this quantity is not currently supported by the mixed model packages, so direct computation based on (6) and (7) is required. The appendix of Hastie (1996) gives some details.

Another useful quantity, also treated in the appendix of Hastie (1996), is the degrees of freedom associated with an estimated function. This is defined as the trace of the matrix that maps the observations to the fitted values and is a decreasing function of the smoothing parameter. Apart from being more interpretable, it has the advantage of being defined for any linear smoother.

5. Smoothing Parameter Selection

For penalized spline model (9), smoothing parameter selection is a by-product of model fitting with variance component estimation. The amount of smoothing for $g(\cdot)$ and $f_\ell(\cdot)$, $\ell=1,\ldots,L$, is governed by

$$\frac{\sigma_\varepsilon^2}{\sigma_{bs}^2} \quad \text{and} \quad \frac{\sigma_\varepsilon^2}{\sigma_{bt}^2+\sigma_{c\ell}^2}, \qquad \ell=1,\ldots,L,$$

respectively. Thus, smoothing parameter selection reduces to variance component estimation in a mixed model, with a small variance component corresponding to more smoothness for a particular curve. Note that models (4) and (9) specify independent amounts of smoothing for each curve f_ℓ. One can obtain either maximum likelihood (ML) or restricted maximum likelihood (REML) estimates (Searle, Casella, and McCulloch, 1992) of the variance components and hence of the smoothing parameters, using, e.g., PROC MIXED in SAS or the S-PLUS function lme. Alternatively, one can fit these models using a prespecified amount of smoothing for a given curve by fixing the value of the corresponding variance component. This can be accomplished, e.g., using the parms option in the SAS procedure PROC MIXED.

Alternatively, one could specify a constant smoothing parameter for the $\{f_\ell\}$ using the simpler model with common variance component $\sigma_{c1}^2=\ldots=\sigma_{cL}^2\equiv\sigma_c^2$. Given model (9), one can test this common smoothness assumption by comparing the appropriate likelihood ratio statistic to a χ_{L-1}^2 variate.

6. Generalized Additive Models with Interactions

The models described in Sections 2 and 3 generalize naturally to the case of nonnormal errors. When the response distribution is a member of the natural exponential family, the relevant penalized spline model can be represented as a generalized linear mixed model (GLMM; Breslow and Clayton, 1993).

Specifically, consider the simple regression setting of Section 2 with triples (y_i,x_i,z_i), but suppose that the responses are independent with density functions of the form

$$f(y_i\mid\theta_i,\phi)=\exp\left[\frac{w_i}{\phi}\{y_i\theta_i-b(\theta_i)\}+c\left(y_i,\frac{\phi}{w_i}\right)\right]. \quad (10)$$

Denote $\mu_i = \mathrm{E}(Y_i)$. A generalized additive model for y_i is

$$g(\mu_i) = f_{z_i}(x_i), \tag{11}$$

where g is a monotone link function. A penalized spline model for (11) is

$$g(\mu_i) = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u}, \tag{12}$$

where $\mathbf{x}_i$ and $\mathbf{z}_i$ are the ith rows of $\mathbf{X}$ and $\mathbf{Z}$, respectively, and $\mathbf{X}$, $\boldsymbol{\beta}$, $\mathbf{Z}$, $\mathbf{u}$, and $\mathbf{G}$ are as defined in Section 2.

Unlike the normal theory case, the marginal likelihood

$$L(\boldsymbol{\beta}, \mathbf{G}; \mathbf{y}) = \int \left[\prod_i f(y_i \mid \mathbf{u}; \boldsymbol{\beta})\right] f(\mathbf{u}; \mathbf{G}) d\mathbf{u}$$

under model (12) is not available in closed form, and one must approximate this integral for maximum likelihood estimation of $\psi = (\boldsymbol{\beta}, \sigma_b^2, \sigma_{c1}^2, \ldots, \sigma_{cL}^2)^{\mathrm{T}}$. Breslow and Clayton (1993) and Wolfinger and O'Connell (1993) proposed pseudo-likelihood (PL) approaches based on Laplace approximation of the integral. The resulting algorithm, which involves iteratively fitting a normal theory linear mixed model to a linearized response, is easy to implement using SAS macro GLIMMIX, and it is the one we use here. The resulting estimates represent approximations to the true maximum likelihood estimates, with more bias for binary responses or large variance components. Alternatively, at the cost of considerable computational effort, one could use Monte Carlo methods, the most popular being Monte Carlo EM (McCulloch, 1997; Booth and Hobert, 1999), to obtain exact ML estimates of ψ. Hobert and Wand (2000) compared Monte Carlo EM to Breslow and Clayton's penalized quasi-likelihood (PQL) approach in the context of nonparametric binary regression and noted that PQL performs well in this context. Extension of these results to semiparametric models is a topic for future research.

For binomial, Poisson, and other distributions falling within the one-parameter exponential family, the parameter ϕ represents an overdispersion parameter (Wolfinger and O'Connell, 1993). For the analysis of the pollen data, we estimate $\hat{\phi}$ using GLIMMIX and rely on informal comparisons of deviances to compare models. An alternative way to account for overdispersion would be to capture variability above that specified by a probabilistic model with additional random effects (Dean, 1992) so that the model reduces to (12) with $\phi = 1$.

7. Analysis of Pollen Data

Stark et al. (1997) and Brumback et al. (2000) used generalized linear and generalized additive models, respectively, to investigate the predictive power of meteorological variables on the daily ragweed pollen counts depicted in Figure 2. The authors were interested in the effects of rain, wind speed, and temperature on daily ragweed pollen count after controlling for seasonal trend in pollen levels.

We now consider both generalized linear and generalized additive interaction models for the pollen data. Let $Y_{ij} \sim \mathrm{Poisson}(\mu_{ij})$ denote the pollen count on day i of pollen season j, $j = 1, \ldots, 4$. We first fit the most general parametric Poisson regression model that contains terms corresponding to rain, wind, temperature, and day in season, with each of these effects varying according to year. Specifically, we fit the Poisson GLM

$$\begin{aligned}\log(\mu_{ij}) = {} & \beta_{0j} + \beta_{1j}x_{ij} + \beta_{2j}w_{ij} + \beta_{3j}t_{ij} + \beta_{4j}r_{ij} + \beta_{5j}i \\ & + \beta_{6j}\log(i+1),\end{aligned} \tag{13}$$

where, for day i in year j, x_{ij} is a rain indicator, w_{ij} denotes wind speed, t_{ij} denotes the fitted values from a smooth of temperature as a function of day in season, and r_{ij} denotes the residual from this smooth. This model corresponds to fitting a Poisson regression model to data from each year separately. This model does not fit the data well, yielding a deviance of 4169.6 on 312 residual d.f.

We next fit generalized additive models to the data to investigate whether this lack of fit arises from the linearity assumptions in the Poisson GLM. Consider the semiparametric regression model

$$\log(\mu_{ij}) = \alpha_0 + \alpha_{1j}x_{ij} + g_1(w_{ij}) + g_2(r_{ij}) + f_j(i). \tag{14}$$

This model specifies a rain-by-year interaction and a factor-by-curve interaction representing distinct seasonal trends for the 4 years. The function forms g_1 and g_2, however, are the same for every year j; i.e., we assume that the relationships between pollen and residual temperature and wind speed do not change from year to year. The PL fit of the appropriate linear penalized spline model yields a deviance of 2577.9 on approximately 293.5 d.f., or an almost 40% decrease relative to the deviance from model (13). Overdispersion is still present, however, and GLIMMIX yields a dispersion parameter estimate of $\hat{\phi} = 8.5$. This overdispersion remains under the more general models containing heterogeneous wind and/or residual temperature effects for different seasons.

Table 1 shows the estimates of the rain coefficients and corresponding standard errors, adjusted for overdispersion, from the fit of model (14). The range of these estimates is larger than those of previous analyses, which is primarily due to the linear assumption for the residual temperature effect in earlier models. Fitting the data to all 4 years simultaneously allows us to investigate the plausibility of a homogeneous rain effect across the 4 years. In particular, we compare the fit of model (14) to that obtained under the constraint $\alpha_{11} = \cdots = \alpha_{14}$. Wolfinger and O'Connell's (1993) PL fit of this simpler model yields a difference in deviance of 2.53 on 3 d.f., suggesting that a common homogeneous rain effect is plausible. Table 1 shows the pooled estimate and associated standard error adjusted for overdispersion. Note the improved precision of the pooled estimate resulting from estimating the rain effect from all of the data.

Figure 3 shows plots of the estimated curves and pointwise 95% confidence bands for the effects of residual temperature

Table 1
Yearly and pooled PL estimates of the rain coefficients from semiparametric model (14)

Year	$\hat{\alpha}_1$	SE
1991	−0.56	0.33
1992	−0.72	0.21
1993	−0.99	0.25
1994	−0.87	0.42
Pooled	−0.80	0.14

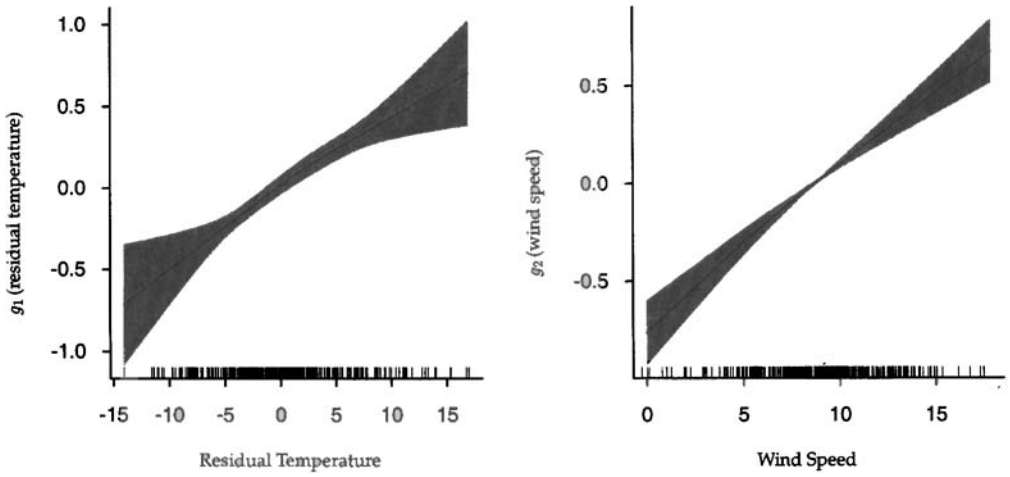

Figure 3. Fitted curves and 95% pointwise confidence bands for the effects of residual temperature and wind speed for the pollen data.

and wind speed on daily pollen counts. Figure 4 shows plots of the estimates and pointwise 95% confidence bands for f_j, $j = 1, \ldots, 4$, from this model. The fit of the model specifying a common rain effect and additivity between year and seasonal trend yields a deviance of 4451.5 on approximately 316.2 residual d.f., supporting the fact that seasonal trend of pollen counts does indeed vary across years.

8. Discussion

In this article, we have proposed a penalized spline approach to estimation of factor-by-curve interactions in additive models. The models have a mixed model representation, which allows the use of existing theory to perform model fitting and smoothing parameter estimation. This approach has the advantage over other smoothing methods in that one does not have to rely on backfitting algorithms. In the pollen example, these methods allowed us to evaluate the plausibility of a common rain effect and common seasonal trends over the four pollen seasons and to gain precision in estimating a common rain effect.

The proposed models specify the functional form of the effect of a continuous covariate X to vary according to the value taken by a categorical predictor Z. However, one can also view the models as specifying the effect of categorical predictor Z on response Y to vary smoothly as a function of

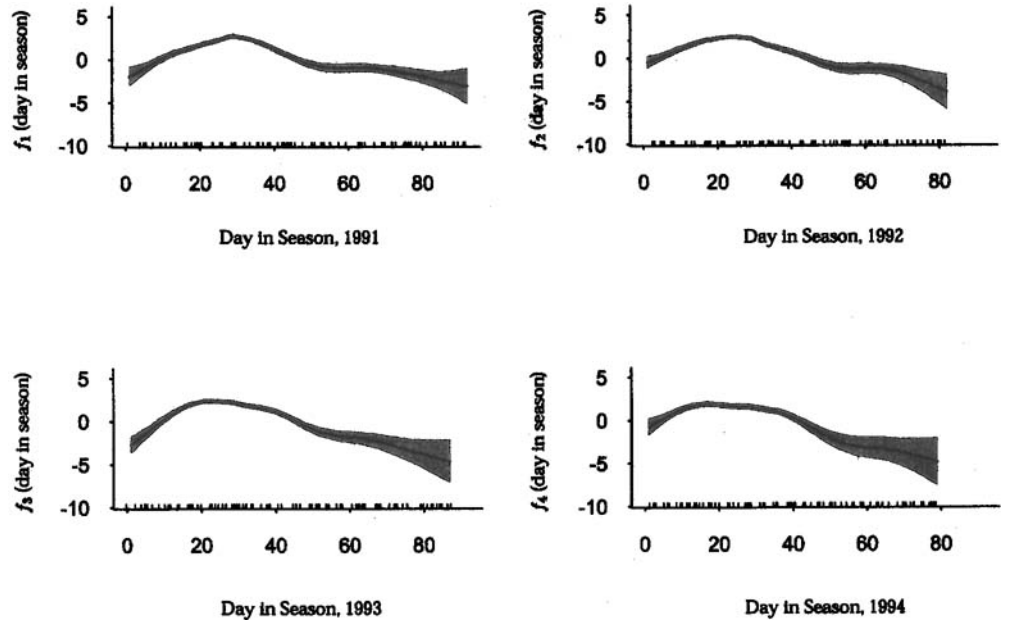

Figure 4. Fitted curves and 95% pointwise confidence bands for the effects of day in season by year for the pollen data.

X. Thus, the models proposed here are discrete analogs of varying-coefficient models (Hastie and Tibshirani, 1993), whereby the functional form of the regression coefficient for Z is estimated using splines.

Because of the conceptual simplicity of the penalized spline approach, extensions to more complex models are straightforward. For instance, one could also use the penalized spline approach to construct bivariate additive interaction models, whereby two-dimensional smooth functions vary across levels of a categorical variable. Finally, because the mixed model framework easily handles correlated data, adaptation of model (5) to the case of dependent errors is straightforward.

Acknowledgements

The authors thank Dr Louise Ryan for encouragement during the early phases of this research and two referees for helpful comments. This research was partially supported by NIH grant ES05860 (BAC), NSF grant DMS 9804058 (DR), and EPA grant R824757 (MPW).

Résumé

La forme fonctionnelle des effets des covariables dans un modèle additif varie souvent entre les groupes définis par les niveaux d'une variable catégorielle. Cette structure caractérise un facteur par interaction courbe. Dans cet article, on présente des modèles de splines avec pénalisation qui incorporent les facteurs par interaction courbe dans les modèles additifs. Une formulation de modèle mixte pour splines avec pénalisation permet un ajustement direct du modèle et la sélection du paramètre de lissage. Nous illustrons le modèle proposé en l'appliquant à des données de pollen d'ambrosie pour lesquelles la tendance saisonnière varie avec l'année.

References

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B* **61,** 265–285.

Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations.* Oxford: Oxford University Press.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Brumback, B. A., Ruppert, D., and Wand, M. P. (1999). Comment to "Variable selection and function estimation in additive nonparametric regression using a data-based prior." *Journal of the American Statistical Association* **94,** 794–797.

Brumback, B. A., Ryan, L. M., Schwartz, J. D., Neas, L. M., Stark, P. C., and Burge, H. A. (2000). Transitional regression models, with application to environmental time series. *Journal of the American Statistical Association* **95,** 16–27.

Chambers, J. M. and Hastie, T. J. (1993). *Statistical Models in S.* London: Chapman and Hall.

Chen, Z. (1987). A stepwise approach for the purely periodic interaction spline model. *Communications in Statistics, Part B, Theory and Methods* **16,** 877–895.

Chen, Z. (1993). Fitting multivariate regression functions by interaction spline models. *Journal of the Royal Statistical Society, Series B* **55,** 473–491.

Coull, B. A., Catalano, P. J., and Godleski, J. J. (2000). Semiparametric analyses of cross-over data with repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics* **5,** 417–429.

Dean, C. B. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association* **87,** 451–457.

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

Hastie, T. J. (1996). Pseudosplines. *Journal of the Royal Statistical Society, Series B* **58,** 379–396.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* London: Chapman and Hall.

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

Hobert, J. P. and Wand, M. P. (2000). Automatic generalized nonparametric regression via maximum likelihood. Technical Report, Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts.

Hobert, J. P., Altman, N. S., and Schofield, C. L. (1997). Analyses of fish species richness with spatial covariate. *Journal of the American Statistical Association* **92,** 846–854.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996). *SAS System for Mixed Models.* Cary, North Carolina: SAS Institute.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92,** 162–170.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects (with discussion). *Statistical Science* **6,** 15–51.

Ruppert, D. and Carroll, R. J. (2000). Spatially-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics* **42,** 205–224.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components.* New York: Wiley.

Stark, P. C., Ryan, L. M., McDonald, J. L., and Burge, H. A. (1997). Using meteorologic data to model and predict daily ragweed pollen levels. *Aerobiologia* **13,** 177–184.

Venables, W. N. and Ripley, B. D. (1994). *Modern Applied Statistics with S-PLUS,* 2nd edition. New York: Springer.

Wahba, G. (1986). Partial interaction spline models for the semiparametric estimation of functions of several variables. *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface,* 75–80.

Wahba, G. (1988). Partial and interaction spline models (with discussion). In *Bayesian Statistics,* Volume 3, J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (eds), 479–491. New York: Oxford University Press.

Wolfinger, R. D. and O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48,** 233–243.

Received January 2000. *Revised October* 2000. *Accepted October* 2000.