Appl. Statist. (2003)
52, *Part* 1, *pp.* 1–18

Geoadditive models

E. E. Kammann and M. P. Wand

Harvard University, Boston, USA

[Received August 2000. Final revision June 2002]

Summary. A study into geographical variability of reproductive health outcomes (e.g. birth weight) in Upper Cape Cod, Massachusetts, USA, benefits from geostatistical mapping or *kriging*. However, also observed are some continuous covariates (e.g. maternal age) that exhibit pronounced non-linear relationships with the response variable. To account for such effects properly we merge kriging with additive models to obtain what we call *geoadditive models*. The merging becomes effortless by expressing both as linear mixed models. The resulting mixed model representation for the geoadditive model allows for fitting and diagnosis using standard methodology and software.

Keywords: Additive models; Disease mapping; Geostatistics; Kriging; Mixed models; Nonparametric regression; Penalized splines; Restricted maximum likelihood

1. Introduction

Geostatistics is concerned with the problem of producing a map of a quantity of interest over a particular geographical region based on, usually noisy, measurements taken at a set of locations in the region. An illustration is provided by Fig. 1. The raw data are longitude, latitude and the residuals from a fitted regression model in which birth weight was regressed against several infant and maternal attributes from an environmental health study in Upper Cape Cod, Massachusetts, USA (see Section 2). The raw data are difficult to visualize and interpret. Fig. 1 is a 'map' of the residuals obtained via the geostatistical method known as *kriging*. It provides an informative summary of the geographical variation in mean birth weight over the region and, in particular, shows possible 'hot spots' of adverse health outcomes. Such hot spots, if found to be significant, are almost inevitably surrogates for unobserved or unknown covariates such as proximity to a source of exposure.

The data that were used to produce Fig. 1 were obtained as part of a study into geographical variation in health outcomes in Upper Cape Cod. Details of the data are given in Section 2. Investigations of this nature are very common and a recent article in *The New Yorker* magazine (Gawande, 1999) reported that, in 1998, the state of Massachusetts responded to more than 3000 disease cluster alarms, most of which concerned cancer. The article was also quite critical of such investigations, pointing out that not one cancer cluster has been convincingly identified. One of the main reasons for this is the lengthy time before the onset of clinical symptoms of many types of cancer. The Upper Cape Cod investigation began as cancer cluster studies but more recently has turned to reproductive outcomes such as birth weight. Reproductive outcomes have the advantage of being more sensitive to recent exposures.

Address for correspondence: M. P. Wand, Department of Biostatistics, School of Public Health, Harvard University, 665 Huntington Avenue, Boston, MA 02115, USA.
E-mail: mwand@hsph.harvard.edu

 0035–9254/03/52001

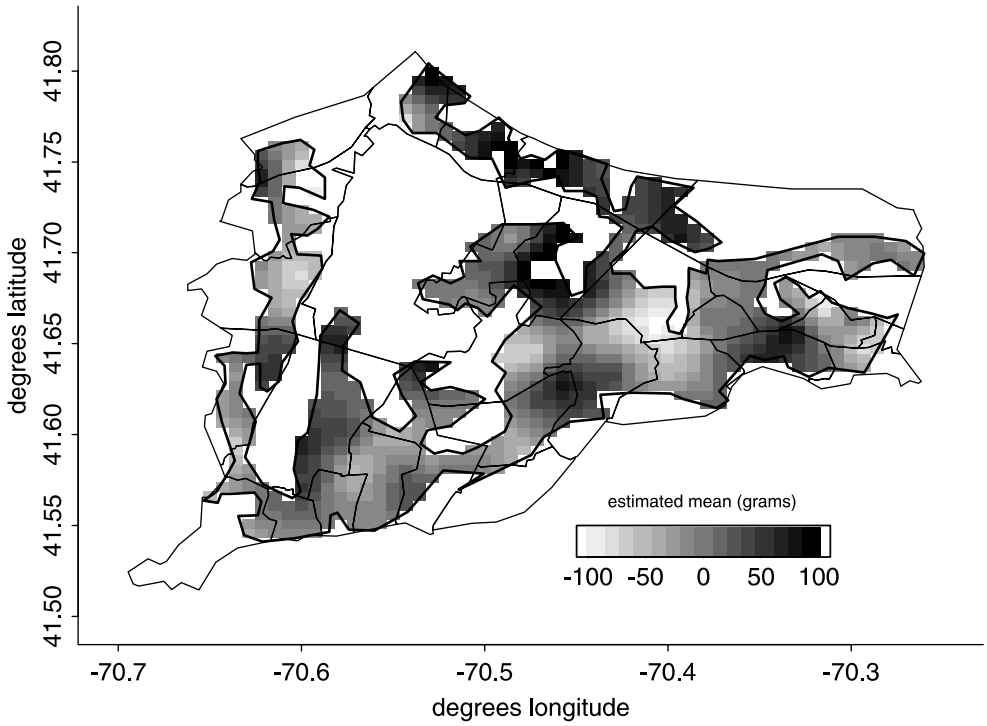

Fig. 1. Map of the residuals from an additive model fit of birth weight to several covariates obtained by using kriging: the data are from the Upper Cape Cod study described in the text

Even for perfect measures of adverse health, kriging alone will not properly address the question of environmental causality. For example, a region with lower income levels is also more likely to have higher levels of adverse health outcomes. The Upper Cape Cod study aims to redress this problem by obtaining data on all other available attributes and accounting for them in the mapping. Fig. 1 represents a cursory attempt to control for covariates. As mentioned above, a regression model was fitted to the attributes and then residuals were mapped. But, ideally, these processes would be done simultaneously. The extension of kriging, sometimes known as *universal kriging* (e.g. Cressie (1993) and Hobert *et al*. (1997)), allows for the incorporation of covariates. However, linearity of the covariate effects is usually assumed. This is not satisfactory for the motivating example since, for instance, maternal age has a non-linear effect on gestational age. Indeed, the regression model that was used to produce Fig. 1 is an *additive* model (e.g. Hastie and Tibshirani (1990)) which permits general smooth functional covariate effects. Our goal is therefore to map reproductive outcomes such as birth weight and gestational age simultaneously while accounting for non-linear covariate effects under the assumption of additivity. The resulting models represent a fusion of geostatistical and additive models: hence the name *geoadditive models*.

There are several ways to combine the ideas of geostatistics and additive modelling. Our research has led to one that has the advantages of being

(a) seamless (because of using a mixed model representation of both kriging and additive models),
(b) model based and likelihood driven (our geoadditive model is simply a linear mixed model and, under Gaussian distributional assumptions, lends itself to an estimation of all

parameters by using (restricted) maximum likelihood and, tentatively at the time of writing, testing via the likelihood ratio paradigm),

(c) low rank, as defined by Hastie (1996) (meaning that the number of basis functions used to construct the function estimates does not grow with the sample size, which is vitally important for disease mapping applications, including the motivating problem, where the data often number in the thousands) and

(d) implementable with standard software (with some simplification in the kriging component we can express the model as a subclass of mixed models commonly known as *variance component models* (e.g. Searle *et al.* (1992)); this leads to enormous reductions in computational complexity and allows for the direct use of standard software such as procedure `MIXED` in SAS and `lme()` in S-PLUS).

Some worthwhile background reading for this paper is Diggle *et al.* (1998) where pure kriging (i.e. no covariates) was the focus. Our paper inherits some of its aspects: model based and with mixed model connections. In particular the comment by Bowman (1998) in the ensuing discussion suggested that additive modelling would be a worthwhile extension. This paper essentially follows this suggestion. However, this paper is not the first to combine the notions of geostatistics and additive modelling. References known to us are Kelsall and Diggle (1998), Durbán Reguera (1998) and Durbán *et al.* (2000). Nevertheless, we believe that our approach has several attractive features (see (a)–(d) above), which are not all shared by these references.

Section 2 describes the motivating application and data in detail. Section 3 shows how we can express additive models as a mixed model and Section 4 does the same for kriging and merges the two into the geoadditive model. Issues concerning the amount of smoothing are discussed in Section 5 and inferential aspects are treated in Section 6. Our analysis of the Upper Cape Cod reproductive data is presented in Section 7. Section 8 discusses an extension to the generalized context. We close the paper with some discussion in Section 9.

2. Description of the application and data

Many environmental health studies have taken place in the region of Massachusetts known as Upper Cape Cod since elevated cancer rates were observed there in the mid-1980s. Several possible sources have been identified and include fuel dumping at a large military reservation, pesticide use in cranberry bogs and polychlorinated biphenyl in water-pipes. However, the studies have been largely inconclusive.

In the late 1990s the Department of Public Health, Commonwealth of Massachusetts, commissioned a new study into the geographical variation of health outcomes in Upper Cape Cod. In the latest phase reproductive outcomes, birth weight and gestational age, have been considered. Birth weight is measured on nearly all newborns and is sensitive to recent exposures, thus facilitating the determination of exposures of biological importance. For example, a decrease of 170–200 g in mean birth weight may be seen in babies whose mothers smoke over 16 cigarettes per day during pregnancy compared with those who do not smoke. Similar arguments can be made for studying gestational age (e.g. National Center for Health Statistics (1996)).

From a statistical viewpoint, birth weight and gestational age have the advantage of being continuous. Fig. 2 gives histograms for these variables corresponding to the Upper Cape Cod data set described below. Apart from the relatively small number of light (e.g. less than 2000 g) or premature (e.g. less than 30 weeks) births both variables are free of any significant skewness. This leads to a simpler model and analysis since the Gaussian assumption is more tenable.

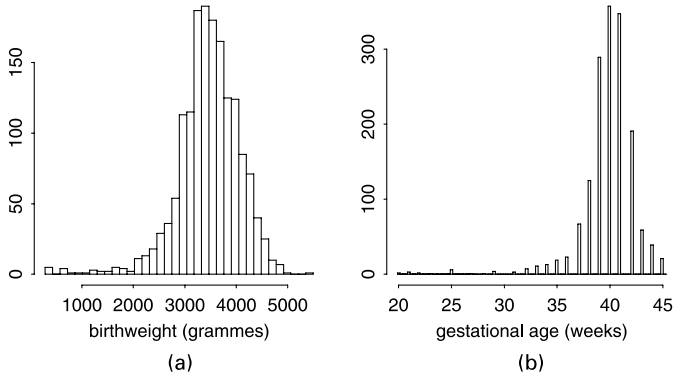

Fig. 2. Histograms of (a) birth weight and (b) gestational age for the Upper Cape Cod reproductive data described in Section 2

Table 1. Covariates that had some association with birth weight and/or with gestational age according to a preliminary analysis, in addition to maternal age, years of education, cigarettes per day and drinks per week†

Covariate	*Description*
Infant covariates	
male	Indicator for infant being male
black	Indicator for infant being black
asian	Indicator for infant being Asian
plurality	$1 \equiv$ single, $2 \equiv$ twin etc.
Maternal covariates	
parity	Number of live-births from the mother
diabetes	Indicator for diabetes
prenatal visits	Number of prenatal care visits
pregnancy hypertension	Pregnancy-related hypertension
incomplete cervix	Indicator for an incomplete cervix
eclampsia	Indicator for eclampsia
light previous birth	Previous preterm infant
heavy previous birth	Previous infant $\geqslant$4000 g
psychiatric	Indicator for psychiatric disorder
renal disease	Indicator for renal disease
uterine bleeding	Indicator for uterine bleeding

†The names are used in the summaries of the analysis in Section 7.

The Upper Cape Cod reproductive data correspond to all 1630 births in 1990 across five towns: Barnstable, Bourne, Falmouth, Mashpee and Sandwich. Apart from geographical location (longitude and latitude) and the outcome variables birth weight (grams) and gestational age (weeks), there are 39 covariates. A preliminary analysis showed that many have no significant association with birth weight or gestational age. Those that are significantly associated with either outcome include maternal age, years of education, number of cigarettes per day and number of alcoholic drinks per week. Table 1 lists all the other variables that exhibited some association with the outcome variables, together with abbreviated names that are used in the summaries of the analysis in Section 7.

3. Penalized spline additive models

The first half of our model formulation involves a low rank mixed model representation of additive models. Mixed model representations in nonparametrics have been used by many researchers in recent years (e.g. Wang (1998), Brumback and Rice (1999), Lin and Zhang (1999) and Verbyla *et al.* (1999)), although early work of this type appears in, for example, Wahba (1978) and Speed (1991). Our approach follows that given by Brumback *et al.* (1999).

For simplicity we shall describe the case of two additive components first. Suppose that (s_i, t_i, y_i), $1 \leqslant i \leqslant n$, represents measurements on two predictors s and t and a response variable y. The additive model for these data is

$$y_i = \beta_0 + f(s_i) + g(t_i) + \varepsilon_i \tag{1}$$

where f and g are smooth, but otherwise unspecified, functions of s and t respectively. Define u_+ to equal u for $u > 0$ and 0 otherwise. A penalized spline version of model (1) involves fitting

$$y_i = \beta_0 + \beta_s s_i + \sum_{k=1}^{K_s} u_k^s (s_i - \kappa_k^s)_+ + \beta_t t_i + \sum_{k=1}^{K_t} u_k^t (t_i - \kappa_k^t)_+ + \varepsilon_i \tag{2}$$

via least squares, but with penalization of the knot coefficients u_k^s and u_k^t (e.g. Marx and Eilers (1998) and Ruppert and Carroll (2000)). Here $\kappa_1^s, \ldots, \kappa_{K_s}^s$ and $\kappa_1^t, \ldots, \kappa_{K_t}^t$ are knots in the s- and t-directions respectively. Rules such as one knot for every three or four unique predictor values, up to a maximum of 20–40 knots, are commonly used, although the sensitivity to this choice is quite low (Ruppert, 2001). A key connection is that penalization of the u_k^s and u_k^t is equivalent to treating them as random effects in a mixed model. Specifically, if we define $\boldsymbol{\beta} = (\beta_0, \beta_s, \beta_t)^{\mathrm{T}}$, $\mathbf{u} = (u_1^s, \ldots, u_{K_s}^s, u_1^t, \ldots, u_{K_t}^t)^{\mathrm{T}}$,

$$\begin{aligned} \mathbf{X} &= (1\ s_i\ t_i)_{1 \leqslant i \leqslant n}, \\ \mathbf{Z} &= (\mathbf{Z}_s | \mathbf{Z}_t) \end{aligned}$$

where

$$\begin{aligned} \mathbf{Z}_s &= ((s_i - \kappa_k^s)_+)_{\substack{1 \leqslant i \leqslant n, \\ 1 \leqslant k \leqslant K_s}} \\ \mathbf{Z}_t &= ((t_i - \kappa_k^t)_+)_{\substack{1 \leqslant i \leqslant n, \\ 1 \leqslant k \leqslant K_t}} \end{aligned} \tag{3}$$

then penalized least squares is equivalent to best linear unbiased prediction in the mixed model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \varepsilon, \qquad E\begin{pmatrix} \mathbf{u} \\ \varepsilon \end{pmatrix} = 0, \qquad \mathrm{cov}\begin{pmatrix} \mathbf{u} \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \sigma_s^2 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_t^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I} \end{pmatrix}. \tag{4}$$

Note that model (4) is a variance components model since the covariance matrix of $(\mathbf{u}^{\mathrm{T}} \varepsilon^{\mathrm{T}})^{\mathrm{T}}$ is diagonal. This is one of the simplest mixed model structures and can be readily fitted by using standard software.

The variance ratio $\sigma_\varepsilon^2 / \sigma_s^2$ acts as a smoothing parameter in the s-direction. Intuitively, a very small value of σ_s^2 leads to overfitting of the truncated lines $(s - \kappa_k)_+$, whereas a very large value leads to a linear fit. Similar comments apply to the t-direction. Smoother fits can be obtained via higher degree spline bases. Alternative representations in terms of B-spline and Demmler–Reinsch bases also exist (Eilers and Marx, 1996; Nychka and Cummins, 1996).

Penalized spline additive models are based on *low rank* smoothers, as defined by Hastie (1996). A precise mathematical definition can be given in terms of the rank of the 'hat' or 'smoother'

matrices, but essentially it corresponds to the number of basis functions staying fixed at $K_s + K_t + 3$, usually about 40–60, regardless of the sample size. For very large n this leads to a computationally less intensive fit with little degradation in the estimator (Hastie, 1996).

The extension to higher numbers of additive components is straightforward. Linear terms are easily incorporated into the model through the $\mathbf{X}\boldsymbol{\beta}$ component. As we shall show in subsequent sections, this mixed model representation has several benefits in terms of model formulation, fitting and diagnosis.

4. Geostatistical extension

Incorporation of a geographical component can be achieved by expressing kriging as a linear mixed model and merging it with an additive model such as model (4) to obtain a single mixed model, which we call the geoadditive model.

Suppose that the data are $(\mathbf{x}_i, y_i), 1 \leqslant i \leqslant n,$ where the y_is are scalar and $\mathbf{x}_i \in \mathbb{R}^2$ represents geographical location. The simple universal kriging model for such data is

$$y_i = \beta_0 + \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x}_i + S(\mathbf{x}_i) + \varepsilon_i \tag{5}$$

where $\{S(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$ is a stationary zero-mean stochastic process and the ε_i are assumed to be independent zero-mean random variables with common variance σ_ε^2 and distributed independently of S (e.g. Cressie (1993)). Prediction at an arbitrary location $\mathbf{x}_0 \in \mathbb{R}^2$ is typically done through an expression of the form

$$\hat{y}(\mathbf{x}_0) = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1^{\mathrm{T}}\mathbf{x}_0 + \hat{S}(\mathbf{x}_0)$$

where $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_1$ are estimates of β_0 and $\boldsymbol{\beta}_1$ respectively and $\hat{S}(\mathbf{x}_0)$ is an empirical best linear unbiased prediction of $S(\mathbf{x}_0)$. For a known covariance structure of S the resulting kriging formula is

$$\hat{y}(\mathbf{x}_0) = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1^{\mathrm{T}}\mathbf{x}_0 + \hat{\mathbf{c}}_0^{\mathrm{T}}(\mathbf{C} + \sigma_\varepsilon^2\mathbf{I})^{-1}(\mathbf{y} - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}_1^{\mathrm{T}}\mathbf{x}_0) \tag{6}$$

where

$$\mathbf{C} = (\mathrm{cov}\{S(\mathbf{x}_i), S(\mathbf{x}_j)\})_{1\leqslant i,j\leqslant n},$$
$$\mathbf{c}_0^{\mathrm{T}} = (\mathrm{cov}\{S(\mathbf{x}_0), S(\mathbf{x}_i)\})_{1\leqslant i\leqslant n}.$$

The practical implementation of equation (6) requires a parsimonious model for the interpoint covariances $\mathrm{cov}\{S(\mathbf{x}), S(\mathbf{x}')\}, \mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$. Following the recommendations of Stein (1999) we use

$$\mathrm{cov}\{S(\mathbf{x}), S(\mathbf{x}')\} = C_{\boldsymbol{\theta}}(\|\mathbf{x} - \mathbf{x}'\|) \tag{7}$$

where $\|\mathbf{v}\| = \sqrt{(\mathbf{v}^{\mathrm{T}}\mathbf{v})}$ and $C_{\boldsymbol{\theta}}$ is a member of the Matérn family of covariance functions. It should be pointed out that equation (7) corresponds to S being *isotropic*, which we view as a reasonable working assumption for the application at hand. The most general such covariance function involves three parameters: $\boldsymbol{\theta} = (\sigma_{\mathbf{x}}^2\, \rho\, \nu)^{\mathrm{T}}$, where $\sigma_{\mathbf{x}}^2 = \mathrm{var}\{S(\mathbf{x})\}$ is the variance of the process, ρ is the *range* parameter and controls the distance at which covariances are effectively zero and ν controls the smoothness of the resulting surface estimate. The full formulation of $C_{\boldsymbol{\theta}}$ is in terms of modified Bessel functions (e.g. Stein (1999), page 31) but the special case $\nu = \frac{3}{2}$ corresponds to

$$C_{\boldsymbol{\theta}}(r) = \sigma_{\mathbf{x}}^2(1 + |r|/\rho)\exp(-|r|/\rho). \tag{8}$$

Indeed, in our analysis we work only with this subfamily of the Matérn covariance functions. We chose equation (8) because it is the simplest member of the Matérn family that results in differentiable surface estimates. We propose to choose ρ via the simple rule

$$\hat{\rho} = \max_{1 \leqslant i, j \leqslant n} \|\mathbf{x}_i - \mathbf{x}_j\| \tag{9}$$

to ensure scale invariance and numerical stability.

These choices of ν and ρ lead to a reduction from three parameters to one to be estimated via restricted maximum likelihood (see Section 5). Apart from the obvious reduction in dimensionality, this also allows for use of standard mixed model software for fitting since kriging reduces to a variance component model (see equation (11) below). Nychka (2000) conjectured that the variance ratio $\sigma_\varepsilon^2/\sigma_{\mathbf{x}}^2$ is much more important than ρ and ν for kriging noisy data. This will be formally investigated in a forthcoming paper by the authors.

Traditionally the $\boldsymbol{\theta}$ in equation (6) is obtained by variogram analysis of the residuals from the detrending fit $\hat{\beta}_0 + \hat{\beta}_1^{\mathrm{T}}\mathbf{x}$, or its quadratic extension, where $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_1$ are chosen via least squares (e.g. Venables and Ripley (1997)). As pointed out by O'Connell and Wolfinger (1997), such an approach is quite *ad hoc*. In addition, Stein (1999) raised concerns about variogram estimation. In keeping with the recommendations of O'Connell and Wolfinger (1997) we propose to use a mixed model approach with residual maximum likelihood for estimation of $\boldsymbol{\theta} = \sigma_{\mathbf{x}}^2$. Precedents of this likelihood approach to kriging include Mardia and Marshall (1984) and Zimmerman (1989).

However, we are still faced with an $n \times n$ matrix inversion. The Upper Cape Cod reproductive data involve $n = 1630$ observations, rendering equation (6) infeasible. An attractive solution is to use *reduced knot or low rank* kriging as proposed by Nychka *et al.* (1998). Let $\{\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_K\}$ be a representative subset of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ which we shall also refer to as knots. This subset can be obtained via an efficient space filling algorithm (e.g. Johnson *et al.* (1990) and Nychka and Saltzman (1998)). Fig. 3 shows the result of applying such an algorithm to the (jittered) locations in the Upper Cape Cod reproductive data. Low rank kriging with space filling algorithms may be viewed as a bivariate extension of low rank scatterplot smoothing as described in Section 3 and, for example, by Hastie (1996) and Eilers and Marx (1996). Earlier references of this type include Parker and Rice (1985), O'Sullivan (1986, 1988), Gray (1992) and Kelly and Rice (1990). Apart from examples in these references, evidence of the effectiveness of low rank smoothing compared with full rank smoothing is provided by French *et al.* (2001), Kammann and Wand (2001) and Ruppert (2001).

Let

$$\mathbf{X} = (1\ \mathbf{x}_i^{\mathrm{T}})_{1 \leqslant i \leqslant n},$$
$$\mathbf{Z} = \Big(C_0(||\,\mathbf{x}_i - \boldsymbol{\kappa}_k\,||/\rho)\Big)_{\substack{1 \leqslant k \leqslant K \\ 1 \leqslant i \leqslant n}}$$

and

$$\boldsymbol{\Omega} = \Big(C_0(||\,\boldsymbol{\kappa}_k - \boldsymbol{\kappa}_{k'}\,||/\rho)\Big)_{1 \leqslant k, k' \leqslant K}.$$

where $C_0(r) = (1 + |r|)\exp(-|r|)$. Then low rank kriging corresponds to fitting the linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{10}$$

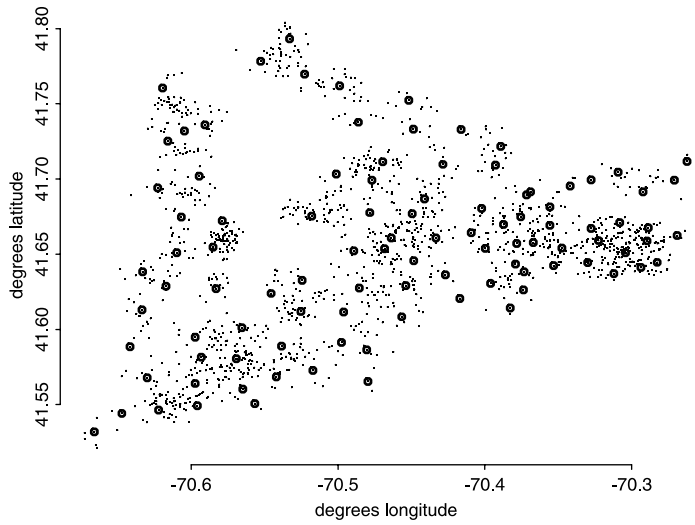

Fig. 3. Geographical locations in the Upper Cape Cod reproductive data (·) with jittering to protect identity, and a representative subset of 100 locations (•) for performing low rank kriging, obtained by using the space filling algorithm of Johnson *et al.* (1990)

where $\text{cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}$ and $\text{cov}(\mathbf{u}) = \sigma_{\mathbf{x}}^2 \boldsymbol{\Omega}^{-1}$. However, for fitting purposes, we should reparameterize to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\tilde{\mathbf{u}} + \boldsymbol{\varepsilon}, \tag{11}$$

where $\tilde{\mathbf{Z}} = \mathbf{Z}\boldsymbol{\Omega}^{-1/2}$ and $\text{cov}(\tilde{\mathbf{u}}) = \sigma_{\mathbf{x}}^2 \mathbf{I}$, and utilize the variance component structure. The best linear unbiased prediction corresponding to equation (10),

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

is nothing more than the set of fitted values on a surface estimate obtained by taking a linear combination of radial basis functions $r_k(\mathbf{x}) = C_0(||\mathbf{x} - \boldsymbol{\kappa}_k||/\rho)$, $1 \leqslant k \leqslant K$, centred near the knots $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_k$. Indeed, it can be viewed as a member of the class of *Matérn splines* described by Handcock *et al.* (1994). The popular surface estimation technique known as thin plate splines (e.g. Wahba (1990) and Green and Silverman (1994)) can also be embedded in this framework through the use of generalized covariance functions. Details are given in French *et al.* (2001). Indeed, the univariate smoothers required for the additive component could also be handled by using such low rank thin plate splines.

The geographical component of an additive model could be handled through other bivariate smoothers such as those based on kernels (e.g. Kelsall and Diggle (1998)). Apart from the advantages of the mixed model representation, we prefer a spline-based approach because of its simpler implementation and better handling of sparse designs.

In view of equations (4) and (11) the geoadditive model

$$y_i = \beta_0 + f(s_i) + g(t_i) + \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x}_i + S(\mathbf{x}_i) + \varepsilon_i \tag{12}$$

is now trivial to formulate as a single linear mixed model. Put

$$\mathbf{X} = (1\ s_i\ t_i\ \mathbf{x}_i^{\mathrm{T}})_{1 \leqslant i \leqslant n},$$

$$\mathbf{Z} = (\mathbf{Z}_s | \mathbf{Z}_t | \mathbf{Z}_{\mathbf{x}})$$

where $\mathbf{Z}_s$ and $\mathbf{Z}_t$ are defined by equations (3) and $\mathbf{Z}_{\mathbf{x}} = \tilde{\mathbf{Z}}$ as given in equation (11). Then the model has representation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{13}$$

where

$$\mathbf{u} = \begin{pmatrix} \mathbf{u}^s \\ \mathbf{u}^t \\ \tilde{\mathbf{u}} \end{pmatrix}, \qquad \operatorname{cov}\begin{pmatrix} \mathbf{u} \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \sigma_s^2\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_t^2\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_{\mathbf{x}}^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_\varepsilon^2\mathbf{I} \end{pmatrix}. \tag{14}$$

This is easily implemented with standard mixed model software. Model (12) can be extended to incorporate linear covariates through the $\mathbf{X}\boldsymbol{\beta}$ term. The extension to more than two additive components is straightforward.

A common convention in additive modelling is to centre the curve estimates about their means. The components of the additive model can be interpreted as effects about the mean. The construction of variability bars (see Fig. 4 later) is more straightforward since they represent the contribution from that term to the overall variability, regardless of the variability in the intercept. The same convention could be applied to the surface estimate in the kriging component of the geoadditive model. Operationally we set $\mathbf{C} = (\mathbf{X}|\mathbf{Z})$ and let $\mathbf{C} = (\mathbf{1}|\mathbf{C}_r)$ be a partition of $\mathbf{C}$ into the intercept column and the remainder. We then work with

$$\bar{\mathbf{C}} = (\mathbf{1}|(\mathbf{I} - (1/n)\mathbf{1}\mathbf{1}^{\mathrm{T}})\mathbf{C}_r) \tag{15}$$

rather than $\mathbf{C}$. This convention is adopted in our analysis in Section 7.

5. Amount of smoothing

Penalized spline regression and kriging are both forms of smoothing and are therefore heavily dependent on the *amount* of smoothing. As mentioned in the previous two sections, the amount of smoothing for both additive components and geostatistical components of a geoadditive model can be quantified through variance component ratios such as $\sigma_\varepsilon^2/\sigma_{\mathbf{x}}^2$. A natural means of choosing the amount of smoothing is to replace variance components with their restricted maximum likelihood estimates (e.g. Searle *et al.* (1992) and O'Connell and Wolfinger (1997)). Since model (14) is a simple variance components model, standard mixed model software such as `PROC MIXED` in SAS or `lme()` in S-PLUS can be called on to obtain a fully automatic fit.

Even in the additive model context, a fully automatic smoothing parameter choice is quite rare. The Markov chain Monte Carlo approaches of Smith and Kohn (1996) and Shively *et al.* (1999) produce automatic additive model fits, and the S-PLUS function `step.gam()` allows for some automation in smoothing-spline-based additive models. However, the more common approach is to use simple rules such as '3 degrees of freedom per additive component' (see Section 5.1), as is the default for the `gam()` function in S-PLUS. Hastie and Tibshirani (1990), pages 159–161, justified this default by arguing that automatic multiple smoothing parameter selection can be somewhat unstable. This is in keeping with work by, for example, Härdle *et al.* (1988) that raises concerns about the instability of automatic smoothing parameter selection even for single-predictor models. Chaudhuri and Marron (1999) recommended looking at curve estimates over a range of smoothing amounts and developed some methodology and graphical devices for doing this systematically.

In summary, although we are attracted by the automatic nature of the mixed model–restricted maximum likelihood approach to fitting geoadditive models, we are reluctant to accept blindly whatever answer it provides, and we recommend looking at other amounts of smoothing.

5.1. Computation of degrees of freedom

We shall now give some details on the computation of degrees-of-freedom values, which are crucial for quantifying the amount of smoothing. For simplicity we restrict our description to model (12). Note that, for example, the degrees of freedom for $f(s)$ are the trace of the matrix that maps the y_is to the $\hat{f}(s_i)$s.

Let $\bar{\mathbf{C}}$ be as defined by equation (15) and let P denote the number of columns in $\bar{\mathbf{C}}$. Then let $\{\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3\}$ be a partition of the column indices $\{1, \ldots, P\}$ such that $\mathcal{I}_0$ corresponds to the intercept β_0, $\mathcal{I}_1$ and $\mathcal{I}_2$ correspond to $f(s)$ and $g(t)$, and $\mathcal{I}_3$ corresponds to $\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x} + S(\mathbf{x})$. For a general matrix $\mathbf{A}$ having P columns define

$$\mathbf{A}_{\mathcal{I}} \equiv \text{submatrix of } \mathbf{A} \text{ consisting of columns with indices in } \mathcal{I}.$$

According to this notation

$$\{\bar{\mathbf{C}}_{\mathcal{I}_0}, \bar{\mathbf{C}}_{\mathcal{I}_1}, \bar{\mathbf{C}}_{\mathcal{I}_2}, \bar{\mathbf{C}}_{\mathcal{I}_3}\}$$

represents a partition of the columns of $\bar{\mathbf{C}}$ corresponding to the terms of the additive model (12). Then the degrees of freedom associated with term j, df_j, can be shown to equal

$$\mathrm{df}_j = \mathrm{tr}[((\bar{\mathbf{C}}^{\mathrm{T}}\bar{\mathbf{C}})_{\mathcal{I}_j})^{\mathrm{T}}\{(\bar{\mathbf{C}}^{\mathrm{T}}\bar{\mathbf{C}} + \sigma_\varepsilon^2\mathbf{B})^{-1}\}_{\mathcal{I}_j}]$$

where

$$\mathbf{B} = \mathrm{diag}\{0, 0, (1/\sigma_s^2)\mathbf{1}_{K_s}, 0, (1/\sigma_t^2)\mathbf{1}_{K_t}, 0, 0, (1/\sigma_{\mathbf{x}}^2)\mathbf{1}_K\}$$

and $\mathbf{1}_p$ denotes a p-dimensional vector of 1s. Although this definition of degrees of freedom is given in the mixed model framework, it matches the definition that is used for ridge regression formulations of penalized splines (e.g. Hastie (1996)).

6. Inference

6.1. Variability bands

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), pages 75–76). Bias aside, they can be interpreted as approximate pointwise confidence intervals (Hastie and Tibshirani, 1990). They are also useful for the detection of leverage and display of inherent variability. For additive models and geoadditive models in the linear mixed model framework the standard errors are easily derived by using standard multivariate statistical manipulations after obtaining an estimate of $\mathrm{cov}\{(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\mathbf{u}}^{\mathrm{T}})^{\mathrm{T}}|\mathbf{u}\}$.

6.2. Hypothesis tests

Another advantage of the mixed model framework is that tests of hypotheses can be performed within the likelihood ratio paradigm. For a general statistical model with data vector $\mathbf{y}$ and parameter vector $\boldsymbol{\theta}$ the test statistic is

$$-2\ \log\{\mathrm{LR}(\mathbf{y})\} = -2\{l(\hat{\boldsymbol{\theta}}_0; \mathbf{y}) - l(\hat{\boldsymbol{\theta}}; \mathbf{y})\} \tag{16}$$

where $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}$ are the (restricted) maximum likelihood estimates of $\boldsymbol{\theta}$ under hypotheses H_0 and H_1 respectively and $l(\boldsymbol{\theta}; \mathbf{y})$ is the log-likelihood. Under the assumption of normal errors equation (16) is easy to compute with standard mixed model software. For example, in model (1), linearity of the effect of s can be assessed through a test of the hypotheses

$$\begin{aligned} H_0 &: \sigma_s^2 = 0, \\ H_1 &: \sigma_s^2 > 0. \end{aligned} \tag{17}$$

The overall effect of s can be assessed through a test of the hypotheses

$$\begin{aligned} H_0 &: \beta_1 = \sigma_s^2 = 0, \\ H_1 &: \beta_1 \neq 0 \text{ or } \sigma_s^2 > 0. \end{aligned} \tag{18}$$

Distribution theory for $-2\ \log\{\mathrm{LR}(\mathbf{y})\}$ under hypothesis H_0 is complicated by the fact that σ_s^2 is on the boundary of its parameter space under H_0 so theory of the type used in Self and Liang (1987) is required. Another complicating factor is the dependence that is induced by the random effects (e.g. Miller (1977)). Even in situations where an asymptotic distributional result has been established for variance components, simulation studies have indicated finite sample discrepancies (e.g. Pinheiro and Bates (2000), page 87). If the dependence is ignored then the theory of Self and Liang (1987) leads to

$$-2\ \log\{\mathrm{LR}(\mathbf{y})\} \xrightarrow{\mathcal{D}} \tfrac{1}{2}\chi_0^2 + \tfrac{1}{2}\chi_1^2 \tag{19}$$

for equation (16) and

$$-2\ \log\{\mathrm{LR}(\mathbf{y})\} \xrightarrow{\mathcal{D}} \tfrac{1}{2}\chi_1^2 + \tfrac{1}{2}\chi_2^2 \tag{20}$$

for test (18). However, Crainiceanu *et al.* (2002) show that the 50:50 mixture of expression (19) does not hold for simple penalized spline models, with considerable differences between restricted and unrestricted maximum likelihood. Other relevant research streams, including one involving the second author, are in progress. Until these investigations have been completed, we are reluctant to report p-values for likelihood ratio tests involving variance components. In the analysis of the Upper Cape Cod reproductive data we shall use expressions (19) and (20) as rough guidelines (see Table 3, later).

7. Analysis of Upper Cape Cod reproductive data

The geoadditive model described in Section 4 was implemented using the S-PLUS function `lme()`, corresponding to version 2.1 of the `NLME` module. The largest geoadditive model required for an analysis of the Upper Cape Cod data took $1\frac{1}{2}$ min to run on our workstations. This is quite fast considering the sophistication of the model and the fact that the smoothing parameter choice is automatic.

We first analysed the data by using a fully automatic smoothing parameter choice based on restricted maximum likelihood. Model selection for the non-linear components was performed using likelihood ratio statistics as described in Section 6.2, whereas the linear components were chosen according to the approximate Z-value given by `lme()`. A full theoretical justification for the use of these Z-values in penalized spline mixed models is being investigated although a cursory justification via, for example, Heckman (1986) is possible. Residuals from final model fits were checked and showed no discernible patterns. Table 2 summarizes the results for the

Table 2. Summary of the final restricted maximum likelihood based fit of the geoadditive model for the Upper Cape Cod reproductive data

Covariate	*Results for birth weight*		*Results for gestational age*	
	Coefficient	*p-value*	*Coefficient*	*p-value*
male	162.80	0.0000		
maternal age	−7.34	0.0067		
pregnancy hypertension	−189.40	0.0472		
light previous birth	−442.20	0.0009		
heavy previous birth	306.60	0.0018		
renal disease	−640.40	0.0551		
black	−148.30	0.0271		
asian	−219.70	0.0517		
drinks per week	−42.34	0.0103		
plurality	−845.30	0.0000	−2.6310	0.0000
uterine bleeding	−412.50	0.0097	−1.3860	0.0418
psychiatric	−525.60	0.0259	−2.0450	0.0430
incomplete cervix	−931.40	0.0485	−3.0750	0.0313
eclampsia	−1074.00	0.0226	−5.4780	0.0066
cigarettes per day			−0.0249	0.0125
	Degrees of freedom			
parity	2.781			
cigarettes per day	2.334			
years of education	3.654			
prenatal visits	2.256		3.331	
maternal age			3.526	
longitude and latitude	2.018		2.003	

model selected. A summary of the likelihood ratio statistics for non-linear effects is given in Table 3. The 90th percentiles of the approximate asymptotic distributions of $-2\log\{\mathrm{LR}(\mathbf{y})\}$ for independent $\mathbf{y}$ (Section 6.2) are included for a rough comparison. In most cases a very high degree of statistical significance is apparent.

Fig. 4 displays all non-linear covariate effects. Although our primary concern in this study is geographical effects on reproductive outcomes, the non-linear covariate effects depicted here are quite interesting in their own right.

Most importantly, the geographical component is *not* found to be significant on the basis of restricted maximum likelihood variance component estimation. As seen in Table 2 restricted maximum likelihood chooses only 2.018 degrees of freedom for location for the prediction of birth weight, and 2.003 degrees of freedom for the prediction of gestational age. This effectively corresponds to a planar fit. The likelihood ratio statistics for non-linearity of the geographical components were both very small. When the model was refitted with longitude and latitude as linear effects the *p*-values were large for both birth weight and gestational age.

Although the analysis based on restricted maximum likelihood showed no geographical effect, we obtained fits where a range of higher degrees-of-freedom values were used for the kriging component. The degrees-of-freedom values for the other non-linear components were fixed at their restricted maximum likelihood values. The results for birth weight are shown in Fig. 5. It suggests some regions with mean birth weights that are lower than average, particularly

Table 3. Likelihood ratio statistics for non-linear terms

Null hypothesis	$-2\ log\{LR(\mathbf{y})\}$ *for birth weight*	$-2\ log\{LR(\mathbf{y})\}$ *for gestational age*
Linearity of parity	6.291	
Linearity of cigarettes per day	3.393	
Linearity of years of education	3.032	
Linearity of prenatal visits	0.741	2.010
Linearity of maternal age		2.440
90th percentile of $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ distribution	1.642	1.642
Effect of parity	31.502	
Effect of cigarettes per day	39.771	
Effect of years of education	21.764	
Effect of prenatal visits		3.810
Effect of maternal age		1.346
90th percentile of $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$ distribution	3.808	3.808

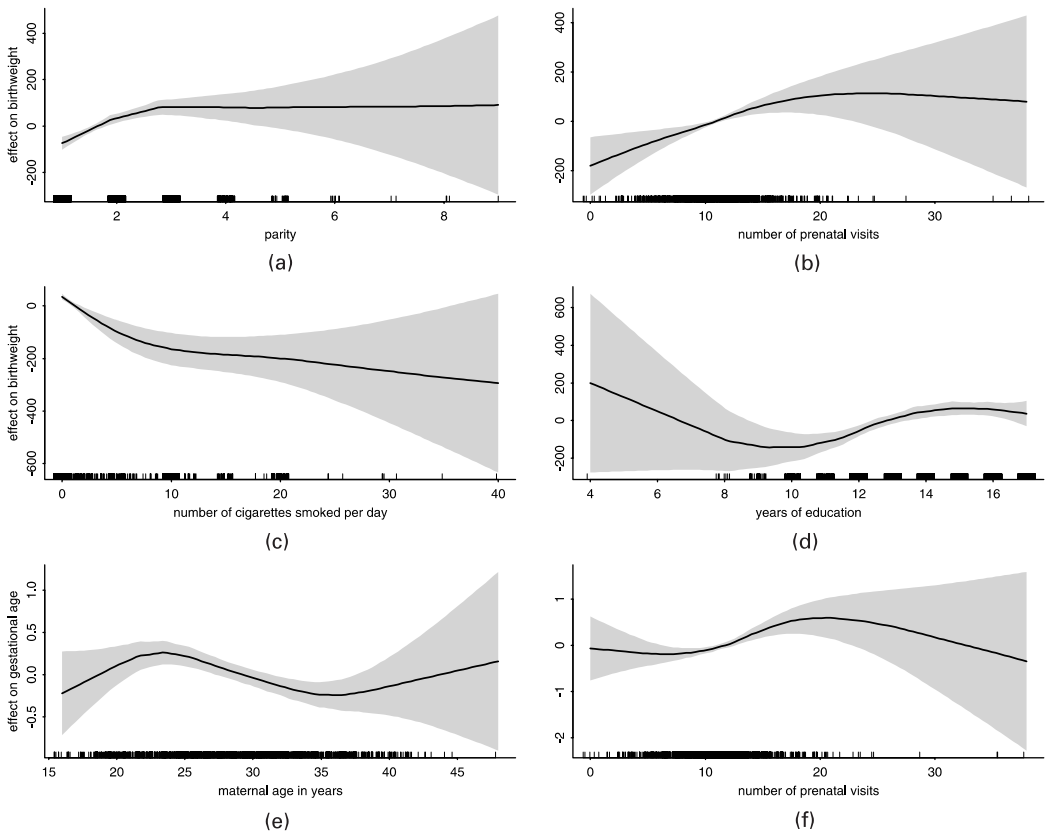

Fig. 4. Non-linear terms from the geoadditive model fit (▒, variability bars correspoiding to ±2 times the estimated standard deviation of the function estimate) for (a)–(d) birth weight and (e), (f) gestational age

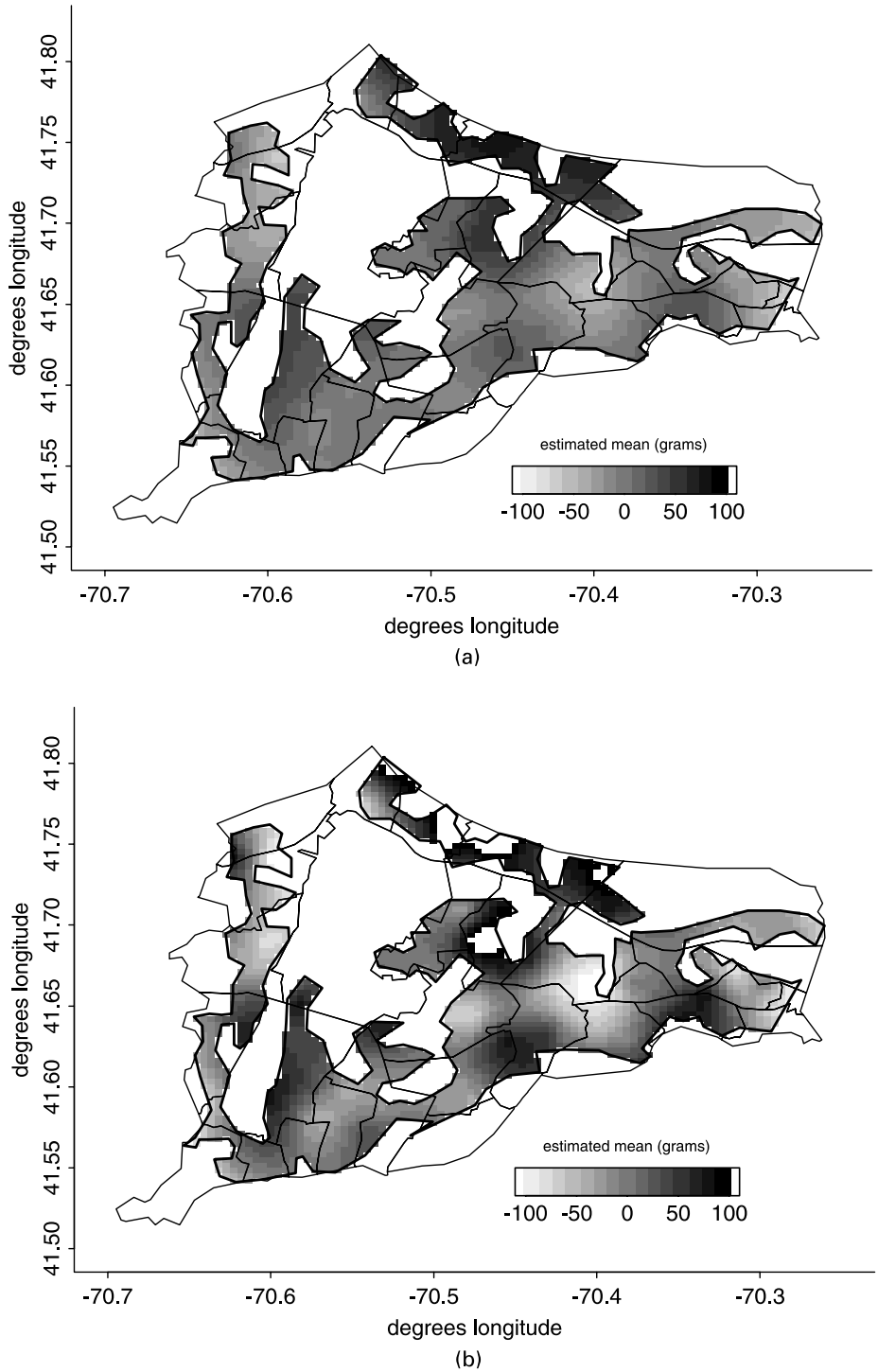

Fig. 5. Geographical components of geoadditive model fits of birth weight with user-specified degrees-of-freedom values (the light lines correspond to census block groups): (a) 20 degrees of freedom; (b) 40 degrees of freedom

in the north-western strip. The Massachusetts Military Reservation is directly east of this strip, and it has long been identified as a source of contamination. Although this result is exploratory, rather than confirmatory, it suggests the possibility of a link between low birth weights and proximity to the military reservation, and it may warrant further investigation.

Fig. 5 is in keeping with the recommendations of Chaudhuri and Marron (1999) who provided some convincing arguments for looking at smoothers across several values of the smoothing parameter, not just that chosen via an automatic method. They developed a graphical device, named *SiZer*, to facilitate the problem of testing for features in a function, while recognizing the inherent dependence on the amount of smoothing in the function estimate. Bivariate extensions have been recently developed (Godtliebsen *et al*., 2000, 2002). An interesting future project would be a SiZer-type analysis of these data to assess systematically the presence of any 'hot spots', after accounting for covariate effects.

An S-PLUS module tailored to fitting geoadditive models has been developed by the authors and is available on request. (The current e-mail address of the second author is `wand@maths.unsw.edu.au`.)

8. Generalized geoadditive models

The reproductive outcomes birth weight and gestational age are continuous and free of any significant skewness, so the Gaussian mixed model is an adequate vehicle for the analysis of those data. In the case where the response is categorical (e.g. a binary or count variable) or heavily skewed, *generalized* linear mixed models need to be used instead. We might call the result *generalized geoadditive models*.

Given the earlier sections, generalized geoadditive models are straightforward to formulate. For example, if the response y is binary then the analogue of equation (12) is

$$\text{logit}\{P(y_i = 1|S)\} = \beta_0 + f(s_i) + g(t_i) + \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x}_i + S(\mathbf{x}_i) \tag{21}$$

and this can be fitted through a mixed model of the form

$$\text{logit}\{P(y_i = 1|\mathbf{u})\} = (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})_i$$

where $\mathbf{u}$ is a random-effects vector with covariance structure exemplified by that given in equations (14). In the case where all covariate effects are linear equation (21) essentially corresponds to the model proposed by Diggle *et al*. (1998).

The fitting of such models by using maximum likelihood is quite complicated owing to intractable integrals in the likelihood. Nevertheless, there has been much research on the topic since the early 1990s (e.g. Breslow and Clayton (1993), Wolfinger and O'Connell (1993), Zeger and Karim (1993), Lin and Breslow (1997), McCulloch (1997), Diggle *et al*. (1998) and Booth and Hobert (1999)) and, in theory, any of these approaches can be used to fit generalized geoadditive models. Future research will investigate the practicalities in the context of geoadditive models.

9. Closing remarks

The geoadditive model is an effective vehicle for the analysis of spatial epidemiologic data and other applications where geographic point data are accompanied by covariate measurements.

The low rank mixed model formulation allows a straightforward implementation and fast processing of large databases, thus facilitating the use of the model in the surveillance of disease clusters.

The geoadditive model has been shown to be useful for an analysis of the Upper Cape Cod reproductive data. It properly accounts for all covariate information before producing disease maps. In the case of gestational age it has been seen that no residual geographical effect is present. The birth weight analysis is slightly suggestive, but geographical variation cannot yet be concluded.

Acknowledgements

We are grateful to Dale Hattis, Bob Knorr, Joel Schwartz, Mike Wright and the Department of Public Health, Commonwealth of Massachusetts, for making the data available for this analysis. The paper has benefited from discussions with Jonathan French, Michael O'Connell, Doug Nychka, Louise Ryan and Lance Waller, and insightful comments from the reviewers. The research of E. E. Kammann was supported by US National Institutes of Health grant T32 ES07142-18.

References

Booth, J. G. and Hobert, J. P. (1999) Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J. R. Statist. Soc.* B, **61**, 265–285.

Bowman, A. W. (1998) Discussion on 'Model-based geostatistics' (by P. J. Diggle, J. A. Tawn and R. A. Moyeed). *Appl. Statist.*, **47**, 334.

Bowman, A. W. and Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis*. Oxford: Clarendon.

Breslow, N. E. and Clayton, D. G. (1993) Approximated inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.

Brumback, B. A. and Rice, J. A. (1999) Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *J. Am. Statist. Ass.*, **93**, 961–994.

Brumback, B. A., Ruppert, D. and Wand, M. P. (1999) Comment on Shively, Kohn and Wood. *J. Am. Statist. Ass.*, **94**, 794–797.

Chaudhuri, P. and Marron, J. S. (1999) SiZer for exploration of structures in curves. *J. Am. Statist. Ass.*, **94**, 807–823.

Crainiceanu, C., Ruppert, D. and Vogelsang, T. (2002) Probability that the MLE of a variance component is zero with applications to likelihood ratio tests. Cornell University, Ithaca. Unpublished.

Cressie, N. (1993) *Statistics for Spatial Data*. New York: Wiley.

Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998) Model-based geostatistics (with discussion). *Appl. Statist.*, **47**, 299–350.

Durbán, M., Hackett, C., Currie, I. and Newton, A. (2000) Analysis of spatial trends in field trials using semiparametric additive models. Unpublished.

Durbán Reguera, M. L. (1998) Modelling spatial trends and local competition effects using semiparametric additive models. *PhD Thesis*. Heriot-Watt University, Edinburgh.

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

French, J. L., Kammann, E. E. and Wand, M. P. (2001) Comment on paper by Ke and Wang. *J. Am. Statist. Ass.*, **96**, 1285–1288.

Gawande, A. (1999) The cancer-cluster myth. *The New Yorker*, Feb. 8th, 34–37.

Godtliebsen, F., Marron, J. S. and Chaudhuri, P. (2000) Significance in scale space. Unpublished.

Godtliebsen, F., Marron, J. S. and Chaudhuri, P. (2002) Significance in scale space for bivariate density estimation. *J. Comput. Graph. Statist.*, **11**, 1–22.

Gray, R. J. (1992) Spline-based tests in survival analysis. *Biometrics*, **50**, 640–652.

Green, P. J. and Silverman, B. W. (1994) *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

Handcock, M. S., Meier, K. and Nychka, D. (1994) Comment on paper by Laslett. *J. Am. Statist. Ass.*, **89**, 401–403.

Härdle, W., Hall, P. and Marron, J. S. (1988) How far are automatically chosen regression smoothing parameters from their optimum? *J. Am. Statist. Ass.*, **83**, 86–101.

Hastie, T. (1996) Pseudosplines. *J. R. Statist. Soc.* B, **58**, 379–396.
Hastie, T. J. and Tibshirani, R. (1990) *Generalized Additive Models*. London: Chapman and Hall.
Heckman, N. E. (1986) Spline smoothing in a partly linear model. *J. R. Statist. Soc.* B, **48**, 244–248.
Hobert, J. P., Altman, N. S. and Schofield, C. L. (1997) Analyses of fish species richness with spatial covariate. *J. Am. Statist. Ass.*, **92**, 846–854.
Johnson, M. E., Moore, L. M. and Ylvisaker, D. (1990) Minimax and maximin distance designs. *J. Statist. Planng Inf.*, **26**, 131–148.
Kammann, E. E. and Wand, M. P. (2001) Low-rank radial smoothing/kriging via restricted maximum likelihood. Harvard University, Boston. Unpublished.
Kelly, C. and Rice, J. (1990) Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics*, **46**, 1071–1085.
Kelsall, J. E. and Diggle, P. J. (1998) Spatial variation in risk of disease: a nonparametric binary regression approach. *Appl. Statist.*, **47**, 559–573.
Lin, X. and Breslow, N. E. (1997) Bias correction in generalized linear mixed models with multiple components of dispersion. *J. Am. Statist. Ass.*, **91**, 1007–1016.
Lin, X. and Zhang, D. (1999) Inference in generalized additive mixed models by using smoothing splines. *J. R. Statist. Soc.* B, **61**, 381–400.
Mardia, K. V. and Marshall, R. J. (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, **72**, 135–146.
Marx, B. D. and Eilers, P. H. C. (1998) Direct generalized additive modeling with penalized likelihood. *Comput. Statist. Data Anal.*, **28**, 193–209.
McCulloch, C. E. (1997) Maximum likelihood algorithms for generalized linear mixed models. *J. Am. Statist. Ass.*, **92**, 162–170.
Miller, J. J. (1977) Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.*, **5**, 746–762.
National Center for Health Statistics (1996) 1990 Birth cohort linked birth/infant death data set. *CD-ROM Series 20, No. 6, SETS Version 1.22a*. National Center for Health Statistics, Hyattsville.
Nychka, D. W. (2000) Spatial process estimates as smoothers. In *Smoothing and Regression* (ed. M. Schimek). Heidelberg: Springer.
Nychka, D. and Cummins, D. J. (1996) Comment on paper by Eilers and Marx. *Statist. Sci.*, **11**, 104–105.
Nychka, D., Haaland, P., O'Connell, M. and Ellner, S. (1998) FUNFITS, data analysis and statistical tools for estimating functions. *Lect. Notes Statist*.
Nychka, D. and Saltzman, N. (1998) Design of air quality monitoring networks. *Lect. Notes Statist.*, **132**, 51–76.
O'Connell, M. A. and Wolfinger, R. D. (1997) Spatial regression models, response surfaces, and process optimization. *J. Comput. Graph. Statist.*, **6**, 224–241.
O'Sullivan, F. (1986) A statistical perspective on ill-posed inverse problems (with discussion). *Statist. Sci.*, **1**, 505–527.
O'Sullivan, F. (1988) Fast computation of fully automated log-density and log-hazard estimators. *SIAM J. Scient. Statist. Comput.*, **9**, 363–379.
Parker, R. L. and Rice, J. A. (1985) Discussion on 'Some aspects of the spline smoothing approach to non-parametric curve fitting' (by B. W. Silverman). *J. R. Statist. Soc.* B, **47**, 40–42.
Pinheiro, J. C. and Bates, D. M. (2000) *Mixed-effects Models in S and S-PLUS*. New York: Springer.
Ruppert, D. (2001) Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *J. Am. Statist. Ass.*, **92**, 1049–1062.
Ruppert, D. and Carroll, R. J. (2000) Spatially-adaptive penalties for spline fitting. *Aust. New Z. J. Statist.*, **42**, 205–224.
Searle, S. R., Casella, G. and McCulloch, C. E. (1992) *Variance Components*. New York: Wiley.
Self, S. G. and Liang, K.-Y. (1997) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Statist. Ass.*, **82**, 605–610.
Shively, T. S., Kohn, R. and Wood, S. (1999) Variable selection and function estimation in additive nonparametric regression using a data-based prior. *J. Am. Statist. Ass.*, **94**, 777–794.
Smith, M. and Kohn, R. (1996) Nonparametric regression using Bayesian variable selection. *J. Econometr.*, **75**, 317–344.
Speed, T. (1991) Comment on paper by Robinson. *Statist. Sci.*, **6**, 42–44.
Stein, M. L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.
Venables, W. N. and Ripley, B. D. (1997) *Modern Applied Statistics with S-PLUS*. New York: Springer.
Verbyla, A. P., Cullis, B. R., Kenward, M. G. and Welham, S. J. (1999) The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Appl. Statist.*, **48**, 269–312.
Wahba, G. (1978) Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. R. Statist. Soc.* B, **40**, 364–372.
Wahba, G. (1990) *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.
Wang, Y. (1998) Mixed effects smoothing spline analysis of variance. *J. R. Statist. Soc.* B, **60**, 159–174.

Wolfinger, R. and O'Connell, M. (1993) Generalized linear mixed models: a pseudo-likelihood approach. *J. Statist. Computn Simuln*, **48**, 233–243.

Zeger, S. L. and Karim, M. R. (1993) Generalized linear models with random effects: a Gibbs sampling approach. *J. Am. Statist. Ass.*, **86**, 79–86.

Zimmerman, D. (1989) Computationally efficient restricted maximum likelihood estimation of generalized covariance functions. *Math. Geol.*, **21**, 655–672.