J. R. Statist. Soc. B (1993)
55, *No.* 2, *pp.* 353–368

Semiparametric Analysis of Variance with Tensor Product Thin Plate Splines

By CHONG GU†

Purdue University, West Lafayette, USA

and

GRACE WAHBA

University of Wisconsin, Madison, USA

[Received March 1991. Revised December 1991]

SUMMARY

We consider fitting multivariate models to response data based on analysis-of-variance (ANOVA)-like decompositions of functions of several variables, $f(t_1, \ldots, t_d) = C + \Sigma_\alpha f_\alpha(t_\alpha) + \Sigma_{\alpha<\beta} f_{\alpha\beta}(t_\alpha, t_\beta) + \ldots$. A theory for fitting (some components of) these models with polynomial smoothing splines exists when each t_α is in a subset of the real line. In this case the various estimated components turn out to be certain tensor sums and products of polynomial splines. This approach may not be natural when one or more of the 'variables' are geographic, in particular where nature does not know north from east. In this case splines of radial structure, such as thin plate splines, for the geographic component, are more natural. In this paper we extend this theory of polynomial smoothing spline ANOVA models to include variables which take values in Euclidean k-space, and fits which turn out to include sums and products of both polynomial and thin plate smoothing splines. The cases of most interest would be $k=2$ and $k=3$. The formulation, interpretation and calculation of the models are discussed, and an application of the technique is illustrated. This work can be used to build predictive ANOVA-like models which describe a response as a function of spatial, temporal and other variables and to explore their interactions. The models can be fitted by using the existing publicly available code RKPACK.

Keywords: MODEL SELECTION; REPRODUCING KERNEL; SPATIAL DATA ANALYSIS; SPLINE ANALYSIS-OF-VARIANCE DECOMPOSITION; TENSOR PRODUCT SMOOTHING SPLINES; THIN PLATE SMOOTHING SPLINES

1. INTRODUCTION

Suppose that we observe

$$y_i = f\{\mathbf{t}(i)\} + \epsilon_i, \qquad i = 1, \ldots, n,$$

where $\mathbf{t} = (t_1, \ldots, t_d)$, $t_\alpha \in \mathscr{T}^{(\alpha)}$, $\alpha = 1, \ldots, d$, and the ϵ_i are independent and identically distributed Gaussian noise with mean 0 and unknown variance σ^2. We want to estimate f from the data. The smoothing spline method estimates f by using the minimizer of

$$\frac{1}{n} \sum_{i=1}^{n} [y_i - f\{\mathbf{t}(i)\}]^2 + \lambda J(f) \tag{1.1}$$

subject to $f \in \mathscr{M}$, where the square error measures the goodness of fit, $J(f)$ is a quadratic roughness penalty, the smoothing parameter λ controls the trade-off of the two and $\mathscr{M}$ is a reproducing kernel Hilbert space (RKHS) of tentative models. See

†*Address for correspondence*: Department of Statistics, Purdue University, West Lafayette, IN 47907, USA.

0035–9246/93/55353 $2.00

Wahba (1990) for an exposition of facts concerning RKHS that we shall be using here. Other references on RKHS are Aronszajn (1950) and Mate (1989).

In this paper we shall assume that f is in an RKHS of smooth real-valued functions on $\mathcal{T} = \mathcal{T}^{(1)} \otimes \ldots \otimes \mathcal{T}^{(d)}$, to be described. $\mathcal{M}$ will actually be composed of tensor sums and products of RK spaces. f will initially be in a large Hilbert space $\mathcal{G}$, in which it has an analysis-of-variance (ANOVA)-like decomposition

$$f(\mathbf{t}) = C + \sum_{\alpha} f_\alpha(t_\alpha) + \sum_{\alpha,\beta} f_{\alpha\beta}(t_\alpha, t_\beta) + \ldots + f_{1,\ldots,d}(t_1, \ldots, t_d). \tag{1.2}$$

Here the f_α are main effects, the $f_{\alpha\beta}$ are the two-factor interactions, etc. We want to exploit the ANOVA structure to effect a compromise between lack of flexibility and the curse of dimensionality by fitting only (some of) the lower order terms in equation (1.2), via a penalized least squares criterion as in expression (1.1). A subspace $\mathcal{M}$ of $\mathcal{G}$ containing only the included terms will be our model space.

There is a growing literature on the special case with $\mathcal{T}^{(\alpha)} = [0, 1]$, $\alpha = 1, \ldots, d$; see Hastie and Tibshirani (1990) and Wahba (1990). Further references are listed at the end of this section.

The f_α, $f_{\alpha\beta}$, etc. are linear combinations of function space marginals of f. In the case $\mathcal{T}^{(\alpha)} = [0, 1]$, these marginals can be obtained by integrating over each variable with respect to the Lebesgue (uniform) measure on [0, 1]. This theory has proved to be very useful for variables t_α whose natural range is an interval of the real line. Main-effect-only models are built into NEW S (Chambers and Hastie, 1991), and additive and interaction smoothing splines may be computed by using RKPACK (Gu, 1989).

In this paper we are interested in extending this theory and practice to where one or more of the predictor variables are spatial, e.g. t_α = (latitude, longitude) for geographical data, or t_α = (latitude–longitude–altitude) for atmospheric data or t_α = (latitude–longitude–depth) for oceanic data. In many examples, the response cannot distinguish north from east, for example, and the tensor product structure of the penalty and interactions terms (to be described) appearing in the existing additive and interaction spline literature is not natural. Then it makes sense to treat these geographic variables in a rotation invariant manner, imposing a thin plate (rotation invariant) structure rather than a tensor product structure on functions of latitude and longitude or latitude, longitude and depth, etc.

It is the purpose of this paper to provide ANOVA spline models that are appropriate when some of the variables are spatial, by extending the additive and interaction spline literature by allowing $\mathcal{T}^{(\alpha)}$ to be $E^{k(\alpha)}$, $k > 1$, and employing thin plate penalty functionals for function components depending on those variables.

This work represents a synthesis and generalization of results in several branches of the statistical and approximation theoretic literature. Some general theory of additive splines is to be found in Stone (1985), Buja *et al.* (1989) and Hastie and Tibshirani (1990). Related references to interaction splines include Barry (1986), Wahba (1986), Chen *et al.* (1989), Friedman and Silverman (1989) and Wahba (1990). Thin plate spline references include Duchon (1977), Meinguet (1979), Wahba and Wendelberger (1980), Utreras (1988) and numerous interesting applications by Hutchinson and collaborators to climatological data; see, for example, Hutchinson *et al.* (1984). Work related to data-based choices of multiple smoothing parameters is found in Gu *et al.* (1989), Gu (1989) and Gu and Wahba (1991a). Here we use the model diagnostics developed by Gu (1992) to select $\mathcal{M}$. A brief preliminary sketch of the main idea here appears in Gu and Wahba (1991b). A one-dimensional special case of the trick that we use to define RKs later goes back at least to deBoor and Lynch (1966). We remark that the field of multivariate function estimation is growing rapidly and in many directions; we just mention two other approaches involving continuous functions, multivariate adaptive regression splines (Friedman, 1991) and the Π-method (Breiman, 1991).

The rest of the paper is organized as follows. In Section 2, we review the existing results for $\mathscr{T}^{(\alpha)} = [0, 1]$ and note some properties of ANOVA in function spaces that will be useful. In Section 3 we present the mathematical result which gives the generalization to the use of thin plate penalties in the ANOVA decomposition. In Section 4, we describe a sample model in detail and show how RKPACK may be used to fit it. An analysis of an application which motivated this research is presented in Section 5. Finally, Section 6 concludes the paper with discussion.

2. ANALYSIS OF VARIANCE IN REPRODUCING KERNEL SPACES: SPECIAL CASE $\mathscr{T}^{(\alpha)} = [0, 1]$

We now give some further background so that the reader can obtain a better idea about what our models are. Returning momentarily to the general case, let $\mathrm{d}\mu_\alpha$ be a probability measure on $\mathscr{T}^{(\alpha)}$, let $\mathscr{H}^{(\alpha)}$ be an RKHS of functions on $\mathscr{T}^{(\alpha)}$ with

$$\int_{\mathscr{T}^{(\alpha)}} f_\alpha(t_\alpha)\,\mathrm{d}\mu_\alpha = 0$$

and let $\{1^{(\alpha)}\}$ be the one-dimensional space of constant functions on $\mathscr{T}^{(\alpha)}$. Letting

$$\mathscr{G} = \prod_\alpha \; [\{1^{(\alpha)}\} \otimes \mathscr{H}^{(\alpha)}],$$

we can expand $\mathscr{G}$ as

$$\mathscr{G} = \{1\} \otimes \sum_\alpha \mathscr{H}^{(\alpha)} \otimes \left[\sum_{\beta<\alpha} \mathscr{H}^{(\alpha)} \otimes \mathscr{H}^{(\beta)} \right] \otimes \ldots . \tag{2.1}$$

Here $\{1\}$ is the space of constant functions on $\mathscr{T}^{(1)} \otimes \ldots \otimes \mathscr{T}^{(d)}$, $f_\alpha \in \mathscr{H}^{(\alpha)}$ is a main effect, $f_{\alpha\beta} \in \mathscr{H}^{(\alpha)} \otimes \mathscr{H}^{(\beta)}$ is a two-factor interaction, and so forth. With some abuse of notation, we are omitting factors of the form $\otimes\{1^{(\alpha)}\}$ wherever they occur in expansion (2.1) multiplied with spaces of non-constant functions. Assuming that the norms on each space $[\{1^{(\alpha)}\} \otimes \mathscr{H}^{(\alpha)}]$ are constructed so that $\{1^{(\alpha)}\} \perp \mathscr{H}^{(\alpha)}$, then the subspaces in square brackets in expansion (2.1) will all be orthogonal in the tensor product norm in $\mathscr{G}$ induced by the original inner products. Once we have selected our model, i.e. decided which subspaces in the right-hand side of expansion (2.1) to include, our model Hilbert space $\mathscr{M}$ would then be the subspace of $\mathscr{G}$ consisting of the direct sum of the selected subspaces. We remark that this description of ANOVA in function spaces is quite general. This idea has also been discussed by del Pino (1991).

In the case $\mathscr{T}^{(\alpha)} = [0, 1]$ described in Wahba (1990), $\mathscr{H}^{(\alpha)} = W_2^m[0, 1] \ominus \{1\}$, where

$$W_2^m[0, 1] \ominus \{1\} = \{f{:}\, f, f', \ldots, f^{(m-1)} \text{ absolutely continuous},$$

$$f^{(m)} \in \mathscr{L}_2[0, 1], \int_0^1 f(t)\,\mathrm{d}t = 0\},$$

$d\mu_\alpha$ has been taken there as uniform measure on [0, 1]. To impose spline penalty functionals, which leave polynomials of degree less than m unpenalized, it is necessary (for $m \geqslant 2$) to decompose the $\mathscr{H}^{(\alpha)}$ further. We shall review how this is done, because in generalizing to the $\mathscr{T}^{(\alpha)}=E^{k(\alpha)}$ case we shall be doing the same thing.

For $\mathscr{T}^{(\alpha)}=[0, 1]$, the polynomial spline penalty functional is

$$J(f) = \int_0^1 f^{(m)2}\,dt. \tag{2.2}$$

Now decompose $\mathscr{H}^{(\alpha)}$ into its unpenalized and penalized parts:

$$\mathscr{H}^{(\alpha)} = \mathscr{H}_\pi^{(\alpha)} \oplus \mathscr{H}_s^{(\alpha)} \tag{2.3}$$

where $\mathscr{H}_\pi^{(\alpha)}$ is spanned by the polynomials of degree less than m constrained to integrate to 0 and $\mathscr{H}_s^{(\alpha)}$ is the 'smooth' subspace. $J(f)$ of equation (2.2) is the norm on $\mathscr{H}_s^{(\alpha)}$, and behind this decomposition there is also a norm defined on $\mathscr{H}_\pi^{(\alpha)}$ and a set of $m-1$ conditions satisfied by the elements of $\mathscr{H}_s^{(\alpha)}$ so that the two subspaces on the right-hand side of equation (2.3) are orthogonal. This norm and set of conditions is not unique and a reasonable choice must be made if interaction terms are to be included in the model. Now, substitute equation (2.3) into equation (2.1) and expand. Thus, for example, $[\mathscr{H}^{(\alpha)} \otimes \mathscr{H}^{(\beta)}]$ becomes

$$[\mathscr{H}^{(\alpha)} \otimes \mathscr{H}^{(\beta)}] = [\mathscr{H}_\pi^{(\alpha)} \otimes \mathscr{H}_\pi^{(\beta)}] \oplus [\mathscr{H}_\pi^{(\alpha)} \otimes \mathscr{H}_s^{(\beta)}] \oplus [\mathscr{H}_s^{(\alpha)} \otimes \mathscr{H}_\pi^{(\beta)}] \oplus [\mathscr{H}_s^{(\alpha)} \otimes \mathscr{H}_s^{(\beta)}]. \tag{2.4}$$

We can, if desired, choose which subspaces to include in the model after this expansion. Then each of the four terms in square brackets on the right-hand side of equation (2.4) is a candidate for leaving in or out, and so forth for the higher order expansions. This allows somewhat more generality than selecting the model (1.2), since terms that are polynomial in one variable and smooth in the the other (second and third terms in equation (2.4)) would then be allowed even if the smooth–smooth term (fourth term in equation (2.4)) is excluded. After that is done all the included subspaces are collected, and the end result is a model space of the form

$$\mathscr{M} = \mathscr{H}^0 \oplus \sum_\beta \mathscr{H}^\beta$$

where $\mathscr{H}^0$ is a finite dimensional space of polynomials which will not be penalized, and the $\mathscr{H}^\beta$ are orthogonal subspaces of the form $\mathscr{H}_s^{(\alpha)}$ or composite spaces of tensor products of two or more spaces of the form $\mathscr{H}_s^{(\alpha)}$ and $\mathscr{H}_\pi^{(\alpha)}$ with at least one of these spaces of the form $\mathscr{H}_s^{(\alpha)}$. The norms on the composite $\mathscr{H}^\beta$ are the tensor product norms induced by the norms on the component subspaces. The estimate f_λ of f is then the minimizer in $\mathscr{M} = \mathscr{H}^0 \oplus \Sigma_\beta \mathscr{H}^\beta$ of

$$\frac{1}{n} \sum_{i=1}^n [y_i - f\{\mathbf{t}(i)\}]^2 + \lambda \sum_\beta \theta_\beta^{-1} \|P^\beta f\|^2, \tag{2.5}$$

where P^β is the orthogonal projector in $\mathscr{M}$ onto $\mathscr{H}^\beta$. An explicit formula for f_λ in the general case given a basis for $\mathscr{H}^0$ and the RKs for each $\mathscr{H}^\beta$ is given in Wahba (1990), and we reproduce it here. Letting ϕ_ν, $\nu=1, \ldots, M$, span $\mathscr{H}^0$ and R_β be the RK for $\mathscr{H}^\beta$, f_λ is given by

$$f_\lambda(\mathbf{t}) = \sum_{\nu=1}^{M} d_\nu \phi_\nu(\mathbf{t}) + \sum_{i=1}^{n} c_i \left[\sum_\beta \theta_\beta R_\beta\{\mathbf{t}(i), \mathbf{t}\} \right]. \tag{2.6}$$

$\mathbf{c} = (c_1, \ldots, c_n)'$ and $\mathbf{d} = (d_1, \ldots, d_M)'$ are solutions to the linear system

$$\left(\sum_\beta \theta_\beta \Sigma_\beta + n\lambda I \right) \mathbf{c} + S\mathbf{d} = \mathbf{y},$$
$$S'\mathbf{c} = 0, \tag{2.7}$$

where Σ_β and S are $n \times n$ and $n \times M$ matrices, defined by $\Sigma_\beta = (R_\beta\{\mathbf{t}(i), \mathbf{t}(j)\})$ and $S = (\phi_\nu\{\mathbf{t}(i)\})$. S must be of full column rank.

Returning to the case $\mathscr{T}^{(\alpha)} = [0, 1]$, RKs for $\mathscr{H}_\pi^{(\alpha)}$ and $\mathscr{H}_s^{(\alpha)}$ (with a particular choice of norm on $\mathscr{H}_\pi^{(\alpha)}$) may be found in Wahba (1990), chapter 10. The RKs for the $\mathscr{H}^\beta$, which are tensor products of spaces of the form $\mathscr{H}_\pi^{(\alpha)}$ and $\mathscr{H}_s^{(\alpha')}$, are then obtained by applying, recursively if necessary, the fact that if $\mathscr{H}^{(1)}$ and $\mathscr{H}^{(2)}$ are RKHSs defined on $\mathscr{T}^{(1)}$ and $\mathscr{T}^{(2)}$, with RKs $R_1(s_1, t_1)$ and $R_2(s_2, t_2)$ respectively, then the RKHS $\mathscr{H}^{(1)} \otimes \mathscr{H}^{(2)}$ has an RK $R(s, t) = R_1(s_1, t_1) R_2(s_2, t_2)$, where $s = (s_1, s_2)$ and $t = (t_1, t_2)$, $s, t \in \mathscr{T} = \mathscr{T}^{(1)} \otimes \mathscr{T}^{(2)}$.

The smoothing parameters $\lambda/\theta_1, \ldots, \lambda/\theta_p$ may be chosen by generalized cross-validation (GCV); see Craven and Wahba (1979) and Wahba (1990). Software for estimating the smoothing parameters by GCV and calculating f_λ via equation (2.6) given the relevant basis functions and RKs is given in the publicly available code RKPACK (Gu, 1989). RKPACK is available by writing to netlib@ornl.gov with the words 'send index'; the robot mailserver will then respond with instructions. Thus, additive and interaction smoothing splines in the $\mathscr{T}^{(\alpha)} = [0, 1]$ case may be fitted by the general public by using RKPACK. In the next section we extend these results to the tensor product thin plate case.

3. ANALYSIS OF VARIANCE WITH THIN PLATE SPLINES

We now proceed to the generalization to the thin plate case. Let $\mathscr{T}^{(\alpha)} = E^{k(\alpha)}$, the Euclidean k-space. The cases of most practical interest are $k=2$ or $k=3$; however, any $k < 2m$ is theoretically possible (provided that there are enough data). The thin plate penalty functional is $J(f) = J_m^k(f)$ given by

$$J_m^k(f) = \sum_{\gamma_1 + \ldots + \gamma_k = m} \frac{m!}{\gamma_1! \ldots \gamma_k!} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \left(\frac{\partial^m f}{\partial x_1^{\gamma_1} \ldots \partial x_k^{\gamma_k}} \right)^2 \mathrm{d}x_1 \ldots \mathrm{d}x_k. \tag{3.1}$$

The penalty J_m^k is invariant under a rotation of the variables $t = t_\alpha = (x_1, \ldots, x_k)$ and as a consequence the thin plate splines are rotation invariant. The thin plate (radial) structure is natural when the variables are qualitatively the same, such as latitude–longitude for geographic data. Frequently in atmospheric data an appropriate scaling of the vertical co-ordinate will allow the same claim to be made for latitude–longitude–altitude.

The fitting of thin plate splines takes place in an abstract RKHS $\mathscr{X}$ (see Meinguet (1979)) of functions for which $J_m^k(f)$ is well defined and finite, and which includes the span of the $M(m, k) = \binom{m+k-1}{k}$ polynomials in k variables of total degree less than m (the null space of J_m^k). The span of these polynomials will play the role of

$[\{1^{(\alpha)}\} \oplus \mathscr{H}_\pi^{(\alpha)}]$ and, once an inner product on all of $\mathscr{X}$ is defined, the orthocomplement of the span of these polynomials in $\mathscr{X}$ will play the role of $\mathscr{H}_s^{(\alpha)}$. $J_m^k(f)$ will be the square norm of $\mathscr{H}_s^{(\alpha)}$. When there is only one 'variable', i.e. $d=1$ and $t=(x_1, \ldots, x_k)$, then the thin plate spline, which is the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - f\{\mathbf{t}(i)\}]^2 + \lambda J_m^k(f), \tag{3.2}$$

is independent of the norm on the null space of J_m^k and can be computed with so-called semikernel or variogram E, to be given below, which is independent of any such norm; see Bates *et al.* (1987). However, to carry out the fitting of the ANOVA model with thin plate components in interaction terms, we need to define a norm on $[\{1^{(\alpha)}\} \oplus \mathscr{H}_\pi^{(\alpha)}]$ and conditions on the functions in $\mathscr{H}_s^{(\alpha)}$ so that $\{1^{(\alpha)}\}$, $\mathscr{H}_\pi^{(\alpha)}$ and $\mathscr{H}_s^{(\alpha)}$ are mutually orthogonal, and we need to provide the RK for $\mathscr{H}_s^{(\alpha)}$. One such RK appears in Wahba and Wendelberger (1980) and it is reprinted (without attribution) in Wahba (1990), section 2.4, where the relation between the semikernel E and the given RK is described in detail. That RK is somewhat artificial for ANOVA use, and since we are not aware of any other published RKs that are suitable for this, we shall now obtain a broad class of RKs (including the aforementioned RK), from which we can select a reasonable RK for use in the ANOVA models that we are proposing.

Let $\mu=\mu_\alpha$ be a probability measure on E^k which assigns probability $p_j > 0$ to the points $u_j \in E^k$, $j=1, \ldots, N$, with the following property. If $\psi_1, \ldots, \psi_M$ are a set of

$$M = M(m, k) = \binom{m+k-1}{k}$$

polynomials which span the null space of J_m^k, then the Gram matrix with (ξ, ν)th entry

$$\langle \psi_\xi, \psi_\nu \rangle_\mu \equiv \sum_{j=1}^{N} p_j \psi_\xi(u_j)\,\psi_\nu(u_j) = \int_{E^k} \psi_\xi \psi_\nu \,\mathrm{d}\mu \tag{3.3}$$

is of full rank M. As an example, for $k=m=2$, this requires that N is at least 3 and not all the u_j fall on a straight line. For the remainder of this section we shall drop the subscripts and superscripts α. If this rank condition is satisfied then equation (3.3) defines an inner product of this space of polynomials. Then we can extract an orthonormal basis, call it $\phi_0, \phi_1, \ldots, \phi_{M-1}$, with $\phi_0(t)=1$, i.e.

$$\langle \phi_\xi, \psi_\nu \rangle_\mu = \delta_{\epsilon,\nu}, \tag{3.4}$$

where $\delta_{\xi,\nu}$ is the Kronecker delta. It follows from the properties of RKHSs that the RK for $\mathscr{H}_\pi = \operatorname{span}\{\phi_1, \ldots, \phi_{M-1}\}$ with the norm defined by equation (3.3) is

$$R_\pi(s, t) = \sum_{\nu=1}^{M-1} \phi_\nu(s)\,\phi_\nu(t). \tag{3.5}$$

To obtain the RKHS for $\mathscr{H}_s$, define, for any $f \in \mathscr{X}$, the projector P_0 by

$$(P_0 f)(t) = \sum_{\nu=0}^{M-1} \phi_\nu(t) \int f\phi_\nu \,\mathrm{d}\mu. \tag{3.6}$$

P_0 will be the orthogonal projection operator in $\mathscr{X}$ onto $\{1\} \oplus \mathscr{H}_\pi$ if we endow $\mathscr{X}$ with the square norm

$$\|f\|^2 = \langle P_0 f, P_0 f\rangle_\mu + J_m^k(f); \tag{3.7}$$

$\mathscr{H}_s$ will be those elements in $\mathscr{X}$ satisfying $P_0 f = 0$. Now, let $E(s, t)$ be the semikernel referred to earlier, which is

$$E(s, t) = \begin{cases} c_m^k \|s-t\|^{2m-k} \log\|2m-k\| & 2m-k \text{ a positive even integer,} \\ c_m^k \|s-t\|^{2m-k} & \text{otherwise} \end{cases} \tag{3.8}$$

where c_m^k are constants given in Appendix A and $\|s-t\|$ is the Euclidean distance between s and t. Let $P_{0(s)}$ be P_0 applied to what follows considered as a function of s.

Theorem 1. The RK R_s for $\mathscr{H}_s$ when $\mathscr{X}$ has the square norm (3.7) is

$$R_s(s, t) = (I - P_{0(s)})(I - P_{0(t)})\, E(s, t). \tag{3.9}$$

A proof is given in Appendix A. The $k = 1$ case is included as a special case when the limits on the integral in equation (2.2) are replaced by $(-\infty, \infty)$. It appears that the result extends to μ any probability measure for which the Gram matrix of equation (3.3) is of full rank and finite, but, except for the $k = 1$ case, generally an approximation with a finite number of mass points would be needed for computations.

4. SAMPLE MODEL

We describe in detail the construction of a sample model for $f(t_1, t_2)$, where $\mathscr{T}^{(1)} = E^1$, the real line, and $\mathscr{T}^{(2)} = E^2$, the Euclidean 2-space. It will be convenient to write $t_2 = (x_1, x_2)$. Taking $m = 2$ for both variables, $\mathscr{H}_\pi^{(1)}$ has as its basis the single linear function $\phi_1^{(1)}$, which satisfies $\int_{E^1} \phi_1^{(1)}\, d\mu_1 = 0$ and $\int_{E^1} \phi_1^{(1)2}\, d\mu_1 = 1$. We have

$$\mathscr{H}_s^{(1)} = \left\{ f: f, f' \text{ absolutely continuous}, f'' \in \mathscr{L}_2, \int_{E^1} f\, d\mu_1 = \int_{E^1} f\phi_1^{(1)}\, d\mu_1 = 0 \right\}.$$

On E^2, $\mathscr{H}_\pi^{(2)}$ has the basis $\{\phi_1^{(2)}, \phi_2^{(2)}\}$, which are linear functions in x_1 and x_2 satisfying

$$\int_{E^2} \phi_1^{(2)}\, d\mu_2 = \int_{E^2} \phi_2^{(2)}\, d\mu_2 = 0$$

and $\int_{E^2} \phi_\mu^{(2)} \phi_\nu^{(2)}\, d\mu_2 = \delta_{\mu,\nu}$, where $\delta_{\mu,\nu}$ is the Kronecker delta. Furthermore,

$$\mathscr{H}_s^{(2)} = \left\{ f, f \in \mathscr{X}, J_m^k(f) < \infty, \int_{E^2} f\, d\mu_2 = \int_{E^2} f\phi_1^{(2)}\, d\mu_2 = \int_{E^2} f\phi_2^{(2)}\, d\mu_2 = 0 \right\}.$$

Turning now to the representation $f \in \mathscr{H}^0 \oplus \Sigma_{\beta=1}^{p} \mathscr{H}^\beta$, $\mathscr{H}^0$ will be spanned by the six elements

$$\{1, \phi_1^{(1)}, \phi_1^{(2)}, \phi_2^{(2)}, \phi_1^{(1)}\phi_1^{(2)}, \phi_1^{(1)}\phi_2^{(2)}\}. \tag{4.1}$$

These six functions on $E_1 \otimes E_2$ become $\phi_1, \ldots, \phi_6$ in equation (2.6). There are $p = 5$ $\mathscr{H}^\beta$s,

$$\{\mathscr{H}_s^{(1)}, \mathscr{H}_s^{(2)}, \mathscr{H}_\pi^{(1)} \otimes \mathscr{H}_s^{(2)}, \mathscr{H}_s^{(1)} \otimes \mathscr{H}_\pi^{(2)}, \mathscr{H}_s^{(1)} \otimes \mathscr{H}_s^{(2)}\}. \tag{4.2}$$

With continued abuse of notation, we are omitting factors $\otimes\{1^{(\alpha)}\}$ in expression (4.2). Strictly, all these subspaces are spaces of functions defined on $E^1 \otimes E^2$.

Now, let us look at the penalties. Considering f as a function defined on $\mathcal{T}^{(1)} \otimes \mathcal{T}^{(2)} = E^1 \otimes E^2$, for $\mathcal{H}^\beta = \mathcal{H}_s^{(1)}$,

$$\|P^\beta f\|^2 = \int_{E^1} \left\{\frac{\partial^2}{\partial t_1^2}\int_{E^2} f(t_1, t_2)\,\mathrm{d}\mu_2\right\}^2 \mathrm{d}t_1.$$

Noting that the factor 1 main effect will be

$$f_1(t_1) = \int_{E^2} f(t_1, t_2)\,\mathrm{d}\mu_2 - \int_{E^1}\int_{E^2} f(t_1, t_2)\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2,$$

we are reassured that $\|P^{\mathcal{H}_s^{(1)}} f\|^2$ amounts to the E^1 spline penalty acting on the main effect. Similarly, for $\mathcal{H}^\beta = \mathcal{H}_s^{(2)}$,

$$\|P^\beta f\|^2 = \int_{E^2} \left\{\left(\frac{\partial^2 f_2}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 f_2}{\partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^2 f_2}{\partial x_2^2}\right)^2\right\} \mathrm{d}x_1\,\mathrm{d}x_2,$$

where $(x_1, x_2) = t_2$ and

$$f_2(t_2) = \int_{E^1} f(t_1, t_2)\,\mathrm{d}\mu_2 - \int_{E^1}\int_{E^2} f(t_1, t_2)\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2$$

is the factor 2 main effect. For $\mathcal{H}^\beta = \mathcal{H}_s^{(1)} \otimes \mathcal{H}_\pi^{(2)}$, we have

$$\|P^\beta f\|^2 = \int_{E^1} \left\{\frac{\partial^2}{\partial t_1^2}\int_{E^2} \phi_1^{(2)}(t_2) f(t_1, t_2)\,\mathrm{d}\mu_2\right\}^2 \mathrm{d}t_1 + \int_{E^1} \left\{\frac{\partial^2}{\partial t_1^2}\int_{E^2} \phi_2^{(2)}(t_2) f(t_1, t_2)\,\mathrm{d}\mu_2\right\}^2 \mathrm{d}t_1$$

and the penalty when $\mathcal{H}^\beta = \mathcal{H}_\pi^{(1)} \otimes \mathcal{H}_s^{(2)}$ is similar. When $\mathcal{H}^\beta = \mathcal{H}_s^{(1)} \otimes \mathcal{H}_s^{(2)}$, we have

$$\|P^\beta f\|^2 = \int_{E^2}\int_{E^2} \left\{\left(\frac{\partial^4 f}{\partial t_1{}^2 \partial x_1^2}\right)^2 + 2\left(\frac{\partial^4 f}{\partial t_1^2 \partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^4 f}{\partial t_1^2 \partial x_2^2}\right)^2\right\} \mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}t_1$$

which of course is different from $J_4^3(f)$.

After the model has been fitted, the various terms which remain in the model are regrouped for display into the mean, the main effect for factor 1, the main effect for factor 2 and the (1, 2) interaction. If no spaces have been deleted, then the respective spaces are mean, $\{1\}$, main effect for factor 1, $\{\phi_1^{(1)}\} \oplus \mathcal{H}_s^{(1)}$, main effect for factor 2, $\{\phi_1^{(2)}\} \oplus \{\phi_2^{(2)}\} \oplus \mathcal{H}_s^{(2)}$, and (1, 2) interaction,

$$\{\phi_1^{(1)}\phi_1^{(2)}\} \oplus \{\phi_1^{(1)}\phi_2^{(2)}\} \oplus [\mathcal{H}_\pi^{(1)} \otimes \mathcal{H}_s^{(2)}] \oplus [\mathcal{H}_s^{(1)} \otimes \mathcal{H}_\pi^{(2)}] \oplus [\mathcal{H}_s^{(1)} \otimes \mathcal{H}_s^{(2)}].$$

In the example that follows we have chosen the μ_α, $\alpha = 1, 2$, as the marginal design measures, i.e. if the data points are $\mathbf{t}(i) = (t_1(i), t_2(i))$, $i = 1, \ldots, n$, then $N = n$ and μ_α on $\mathcal{T}^{(\alpha)}$ assigns mass n^{-1} to the points $(t_\alpha(1), \ldots, t_\alpha(n))$. The μ_α can be chosen independently of the design if there is reason for doing so. When $\mathcal{T}$ is the unit d-cube then the Lebesgue measure in each variable is natural, but this does not generalize in a natural way to the thin plate spline case where the data may be scattered in irregular shapes. For $k = 2$ or $k = 3$, if it were desired to use a continuous density, most probably a quadrature formula would have to be used to compute the required RKs, which would essentially be equivalent to using an approximating measure with a finite number of mass points.

Given measures μ_α with a finite number of mass points and a convenient basis $\psi_0=1, \psi_1^{(\alpha)}, \ldots, \psi_{M-1}^{(\alpha)}$ for the null space of $J_m^k=J_{m(\alpha)}^{k(\alpha)}$, an orthonormal basis $\phi_0=1, \phi_1^{(\alpha)}, \ldots, \phi_{M-1}^{(\alpha)}$ in the sense of equation (3.4) can be obtained by the *QR*-decomposition without pivoting. When the μ_α are the marginal design measures, the formulae become relatively simple. For example, letting S_α be the $n \times M$ matrix with $i\nu$th entry $\phi_\nu^{(\alpha)}\{t_\alpha(i)\}$ and E_α be the $n \times n$ matrix with ijth entry $E\{t_\alpha(i), t_\alpha(j)\}$, then the matrix Σ_β appearing in equations (2.7) which corresponds to $\mathscr{H}^\beta=\mathscr{H}_s^{(\alpha)}$ is, by substituting into theorem 1, given by $(I-S_\alpha S_\alpha'/n)E_\alpha(I-S_\alpha S_\alpha'/n)$. Corresponding matrices for the tensor product spaces are obtained by the componentwise multiplication described in the discussion at the end of the paragraph following equations (2.7).

5. EXAMPLE

In this section, we analyse some environmental data which motivated our conception of tensor product thin plate splines. The data that we shall be analysing are derived by Douglas and Delampady (1990) from the Eastern Lake Survey of 1984 implemented by the Environmental Protection Agency of the USA. The derived data set contains geographic information, water acidity measurements and main ion concentrations of 1798 lakes in three regions, north-east, upper midwest and south-east, in eastern USA. Of interest is the dependence of the water acidity on the geographic locations and other information concerning the lakes. Preliminary analysis and consultation with a water chemist suggested that a model for the surface pH in terms of the geographic location and the calcium concentration is appropriate. As illustrations of the methodology, we only present analysis for lakes in the south-east, which can be further divided into two disconnected subregions—Blue Ridge with 112 lakes and Florida with 159 lakes.

We used the sample model of Section 4 with the μ_α as the marginal design measures. t_1 was taken as the logarithm of the calcium concentration (milligrams per litre). $t_2=(x_1, x_2)$ were obtained by converting the longitude and the latitude of the lake location to the east–west and north–south distances from a local centre. Since the calcium concentration and (latitude, longitude) each have their own smoothing parameters we do not have to worry about the relative scale of these variables, since individual scales are absorbed into the smoothing parameters, which are estimated from the data.

The calculations of the models were performed by using the generic algorithms of Gu and Wahba (1991a) which compute f_λ from equation (2.6) and estimate the $\lambda/\theta_1, \ldots, \lambda/\theta_p$ by GCV. These algorithms are implemented in RKPACK (Gu, 1989), with $\mathscr{H}^0$ as in expression (4.1) and the five $\mathscr{H}^\beta$s as in expression (4.2). The RKs R_β are built up from equations (3.5) and (3.9).

Evaluating a computed model at the design points, we derive a retrospective linear model $\mathbf{y} = \tilde{\mathbf{f}}_0 + \tilde{\mathbf{f}}_1 + \tilde{\mathbf{f}}_2 + \tilde{\mathbf{f}}_{1,2} + \tilde{\mathbf{e}}$, and adjusting for the constant effect by projecting each component onto $\{\mathbf{1}\}^\perp$ we obtain $\mathbf{z} = \mathbf{f}_1 + \mathbf{f}_2 + \mathbf{f}_{1,2} + \mathbf{e}$. To measure the concurvity in the fit, we use the collinearity indices of Stewart (1987), $\kappa_i = \|\mathbf{f}_i\|\,\|\mathbf{f}_i^{(+)}\|$ where $\mathbf{f}_i^{(+)}$ is the ith row of the Moore–Penrose inverse of $(\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_{1,2})$, which can be computed from $\cos(\mathbf{f}_i, \mathbf{f}_j)$. The **f**s are supposed to predict the 'response' **z** so a nearly orthogonal angle between an $\mathbf{f}_i$ and **z** indicates a noise term. Signal terms should be reasonably orthogonal to the residuals; hence a large cosine between an $\mathbf{f}_i$ and **e** makes

TABLE 1
Diagnostics for the Florida model†

	$\mathbf{f}_1$	$\mathbf{f}_2$	$\mathbf{f}_{1,2}$	$\mathbf{e}$	$\mathbf{z}$
κ	1.07	1.13	1.11		
cos(**z**,)	0.861	0.045	0.076	0.457	1
cos(**e**,)	0.007	0.106	0.129	1	0.457
‖ ‖	14.53	2.62	2.23	6.53	15.77

†$R^2 = 0.793$.

TABLE 2
Diagnostics for the Blue Ridge model†

	$\mathbf{f}_1$	$\mathbf{f}_2$	$\mathbf{f}_{1,2}$	$\mathbf{e}$	$\mathbf{z}$
κ	1.08	1.07	1.03		
cos(**z**,)	0.648	0.574	0.358	0.617	1
cos(**e**,)	0.000	0.124	0.249	1	0.617
‖ ‖	2.45	1.44	1.14	2.07	4.10

†$R^2 = 0.632$.

a term suspect. cos(**z**, **e**) and $R^2 = \|\mathbf{z}-\mathbf{e}\|^2/\|\mathbf{z}\|^2$ are informative *ad hoc* measures for the signal-to-noise ratio in the data. A *very* small norm of an $\mathbf{f}_i$ relative to that of **z** also indicates a negligible term. More discussion of these diagnostics can be found in Gu (1992).

For the Florida model, the diagnostics are summarized in Table 1. The diagnostics indicate that the interaction and the geography main effect are absent. Fitting a standard cubic smoothing spline in t_1 gives the final model plotted in Fig. 1, where a scatterplot of the data is superimposed. We concluded that the water acidity of the lakes surveyed in Florida did not illustrate any spatial pattern other than uniformity.

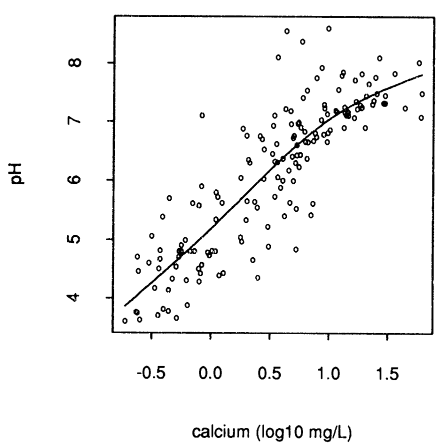

Fig. 1. Final Florida model

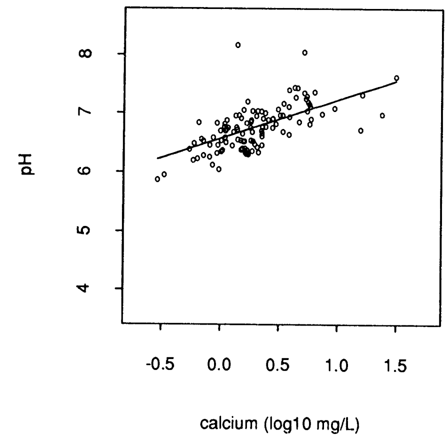

Fig. 2. Calcium main effect for the Blue Ridge model

For the Blue Ridge model the diagnostics are summarized in Table 2. All three components seem non-negligible. The calcium main effect is plotted in Fig. 2. The geography main effect is plotted in Fig. 3. We plotted contours only where there are data. We have done this free hand, but we are developing a method which should give an objective measure of how far one can reasonably extend the model beyond the data. The crest of the southern Blue Ridge mountains runs roughly from south-west to north-east with the highest peak slightly to the left of the centre of Fig. 3. Since geography could be just a proxy of elevation, we also tried to fit a model on calcium and elevation. The diagnostics of such a fit are summarized in Table 3. Again we would preserve all three terms but the diagnostics suggested that elevation could not approximately replace geography in modelling pH. We finally plotted the residuals from the calcium-geography model *versus* the residuals from the calcium-elevation model in Fig. 4, to illustrate the spread of the residuals and to double-check that the calcium-geography model was indeed a refinement of the calcium-elevation model.

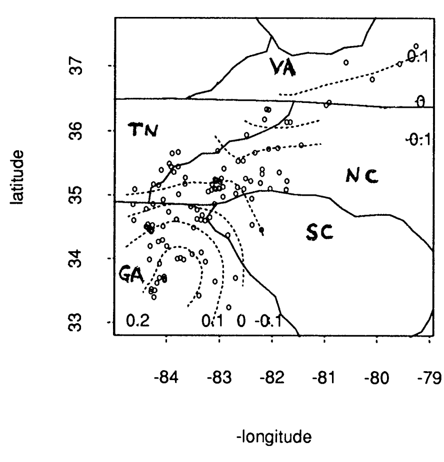

Fig. 3. Geography main effect for the Blue Ridge model

TABLE 3
Diagnostics for the calcium-elevation model†

	f_1	f_2	$f_{1,2}$	e	z
κ	1.40	1.17	1.22		
cos(z,)	0.714	0.371	0.375	0.713	1
cos(e,)	0.038	0.000	0.075	1	0.713
‖ ‖	2.50	0.59	0.26	2.81	4.10

†$R^2=0.493$.

6. DISCUSSION

The models that we have constructed in this paper provide convenient non-parametric tools for combining spatial patterns with other structures of the data. Our examples illustrate how to combine geography with another continuous factor of unknown form. To incorporate an additive term of known form, the partial spline structure (Wahba, 1986) can be adopted. For example, if we wish to include, say, a watershed factor, we simply include a term $x\beta$ in the Blue Ridge pH model where $x=\pm 1$ depending on whether the lake is located on the inland side or the ocean side of the ridge (i.e. x is an indicator function for two watersheds). In dealing with three geographic variables, e.g. longitude, latitude and ocean depth, we might still want a thin plate main location effect, with depth rescaled by changing the units for depth by a multiplier. In principle, this multiplier can also be chosen by GCV; see Hutchinson *et al.* (1984). Similar remarks may apply to time. Replacing mean-square goodness of fit by mean negative log-likelihood, the same structure can be used to fit odds ratios for binary data and to fit intensities for Poisson data, etc.; see for example, Gu (1990). Other extensions include models with inequality constraints, models based on aggregated data, etc.: see Wahba (1990).

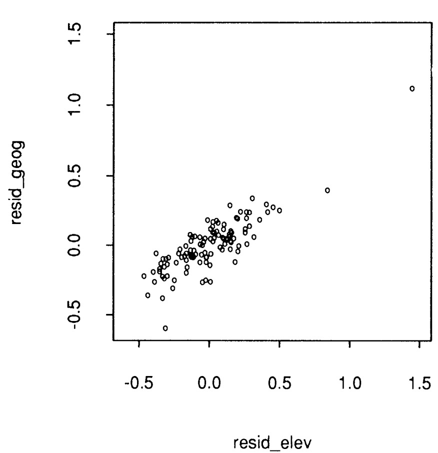

Fig. 4. Residuals for the Blue Ridge models

Compared with model construction, model selection is much less well understood. We have chosen to use the model selection procedures based on a retrospective linear model given in Gu (1990a). Although the theory does not exist for determining when a $\cos(\mathbf{z}, \mathbf{f}_i)$ term is small in the context of nonparametric regression, clearly these diagnostics are informative and could be calibrated intuitively. We are investigating another approach to model selection in the same spirit, based on the Bayesian 'confidence intervals' discussed by Wahba (1983) and Nychka (1988). In principle, the GCV function V (see Wahba (1990) for the definition) could be used for model building. In fact, if the estimate of a θ corresponding to a subspace is 0, then that subspace is automatically removed. However, the GCV is a predictive criterion, and there appears to be a moderately large chance that the estimated θ_β is not 0 even when the true f has $\|P^\beta f\| = 0$ (Wahba, 1990). Thus small components generated by noise may be retained. We could, in principle, develop a hypothesis-testing approach for deciding when θ_β is *significantly* non-zero, by looking at the statistic $v = V(. . .; \infty)/V(. . .; \lambda\hat{\theta}_\beta^{-1})$. Similarly, a likelihood ratio statistic $m = M(. . .; \infty)/M(. . .; \lambda\hat{\theta}_\beta^{-1})$ (see Wahba (1990) for the definition) could be used if the Bayesian model associated with smoothing splines is true. Major practical problems remain, however, in generating distributions for these tests, by Monte Carlo or other methods. Asymptotic χ^2-distributions generally will not be useful because the null model is on the *boundary* of the parameter space of the θs. Tests of the null hypothesis that the true model is in $\mathscr{H}^0$ are discussed in Cox *et al.* (1988) and Wahba (1990). Distributions under this (simple) null hypothesis are relatively straightforward to obtain by Monte Carlo methods.

ACKNOWLEDGEMENT

This research was supported by the US National Science Foundation under grants DMS-9002566 and DMS-9101730.

APPENDIX A: PROOF OF THEOREM 1

To show that $R(s, t)$ is the RK for a given space, we need to show that $R(t, \ )$ is in the space, for each fixed t, and that $\langle R(s, \ ), R(t, \ )\rangle = R(s, t)$. The theorem will follow from a result of Meinguet (1979), which we give after a few definitions. Given m and k, a set of $N+1$ points in R^k is said to be unisolvent, if least squares regression on $\phi_0, . . ., \phi_{M-1}$, the null space of J_m^k, is unique. A set of (unisolvent) points $u_0, u_1, . . ., u_N$ and associated weights $h_0, h_1, . . ., h_N$ is called a generalized divided difference if

$$\sum_{k=0}^{N} h_k \phi_\nu(u_k) = 0, \qquad \nu = 0, 1, . . ., M-1.$$

Let

$$\langle f, g\rangle_* = \sum_{\gamma_1+\ldots+\gamma_k=m} \frac{m!}{\gamma_1! \ldots \gamma_d!} \int \ldots \int \frac{\partial^m f}{\partial x_1^{\gamma_1} \ldots \partial x_k^{\gamma_k}} \frac{\partial^m g}{\partial x_1^{\gamma_1} \ldots \partial x_k^{\gamma_k}} \mathrm{d}x_1 \ldots \mathrm{d}x_k. \tag{A.1}$$

We have the following theorem.

Theorem 2. (Meinguet, 1979). Let $\{h_k, u_k\}$ be a generalized divided difference, and let $E(s, t)$ be defined as in equations (3.8). Let $E_t(\ ) = E(t, \ )$. Then $\Sigma_{j=0}^N h_j E_{u_j} \in \mathscr{X}$ and

$$\left\langle \sum_{j=0}^{N} h_j E_{u_j}(\ ),\ \sum_{k=0}^{N} h_k E_{u_k}(\ ) \right\rangle_* = \sum_j \sum_k h_j h_k E(u_j, u_k).$$

We can now prove theorem 1. We have

$$R(t,\ ) = E_t(\ ) - \sum_{\nu=0}^{M-1} \phi_\nu(t) \sum_{k=0}^{N} p_k \phi_\nu(u_k) E_{u_k}(\ ) + \pi(\ ),$$

where $\pi(\ )$ is a polynomial of total degree less than m. Rewrite this as

$$E_t - \sum_{k=1}^{N} h_k(t) E_{u_k}(\ ) + \pi(\ ),$$

where

$$h_k(t) = \sum_{\nu=0}^{M-1} \phi_\nu(t) p_k \phi_\nu(u_k).$$

Now, we show that $(1, -h_1(t), \ldots, -h_N(t); t, u_1, \ldots, u_N)$ is a generalized divided difference for every t. To do this we have to show that

$$\phi_\nu(t) - \sum_{k=1}^{N} h_k(t) \phi_\nu(u_k) = 0, \qquad \mu = 0, \ldots, M-1. \tag{A.2}$$

Substituting the definition of h_k into the left-hand side of equation (A.2) gives

$$\phi_\nu(t) - \sum_{\xi=0}^{m-1} \phi_\xi(t) \langle \phi_\xi, \phi_\nu \rangle_\mu \tag{A.3}$$

which equals 0 for each ν by the orthonormality of the $\{\phi_\nu\}$ under $\langle\ ,\ \rangle_\mu$. It then follows from Meinguet's theorem that $\langle R(t,\ ), R(s,\ )\rangle_* = R(s, t)$. Since $P_{0(s)}$ is idempotent it follows that $P_{0(s)} R(t, s) = 0$ for all t, so $R(t,\ )$ is in $\mathscr{H}_s = \mathscr{H}_s^{(\alpha)}$, and we have finished the proof.

For completeness, we give the constants

$$c_m^k = \frac{(-1)^{k/2+m+1}}{2^{2m-1} \pi^{k/2} (m-1)!(m-k/2)!}, \qquad k \text{ even},$$

$$c_m^k = \frac{\Gamma(k/2-m)}{2^{2m} \pi^{k/2} (m-1)!}, \qquad k \text{ odd}.$$

If they are omitted in the calculations they will be absorbed by the smoothing parameters.

We remark that the construction here involving polynomials of total degree less than m on R^k will work if E is replaced by *any* function F which is m conditionally positive definite; see Micchelli (1986). F is m conditionally positive definite if $\Sigma_{j,k} h_j h_k F(s, t) \geqslant 0$ whenever $\{h_k; u_k\}$ is a generalized divided difference with respect to the polynomials of total degree less than m. This condition is exactly what we need to show that $(I - P_{0(t)})(I - P_{0(s)}) F(s, t)$ is a non-negative definite function, i.e. an RK. We could also replace $E(s, t)$ by $F(s, t) = \exp\{-\theta(\|t-s\|)^2\}$, which has been used by Sacks *et al.* (1989) and others, or any other positive definite function, with any m for which S_α is of full column rank. The polynomials can be replaced by other functions under certain conditions; see Kimeldorf and Wahba (1971), Wahba (1978) and Ramsay and Dalzell (1991). See Wahba (1990), chapter 3, for a discussion on limiting the class of models that one would want to consider.

REFERENCES

Aronszajn, N. (1950) Theory of reproducing kernels. *Trans. Am. Math. Soc.*, **68**, 337-404.

Barry, D. (1986) Nonparametric Bayesian regression. *Ann. Statist.*, **14**, 934-953.

Bates, D., Lindstrom, M., Wahba, G. and Yandell, B. (1987) GCVPACK—routines for generalized cross validation. *Communs Statist. Simuln Computn*, **16**, 263-297.

deBoor, C. and Lynch, R. E. (1966) On splines and their minimum properties. *J. Math. Mech.*, **15**, 953-969.

Breiman, L. (1991) The Π method for estimating multivariate functions from noisy data. *Technometrics*, **33**, 125-160.

Buja, A., Hastie, T. and Tibshirani, R. (1989) Linear smoothers and additive models. *Ann. Statist.*, **17**, 453-555.

Chambers, J. and Hastie, T. (1991) *Statistical Models in S*. Pacific Grove: Wadsworth and Brooks/Cole.

Chen, Z., Gu, C. and Wahba, G. (1989) Discussion of "Linear smoothers and additive models" by Buja, Hastie and Tibshirani. *Ann. Statist.*, **17**, 515-521.

Cox, D. D., Koh, E., Wahba, G. and Yandell, B. (1988) Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.*, **16**, 113-119.

Craven, P. and Wahba, G. (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**, 377-403.

Douglas, A. and Delampady, M. (1990) Eastern lake survey—phase I: documentation for the data base and the derived data sets. *SIMS Technical Report 160*. Department of Statistics, University of British Columbia, Vancouver.

Duchon, J. (1977) Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables*, pp. 85-100. New York: Springer.

Friedman, J. (1991) Multivariate adaptive regression splines. *Ann. Statist.*, **19**, 1-141.

Friedman, J. and Silverman, B. (1989) Flexible parsimonious smoothing and additive modeling. *Technometrics*, **31**, 3-21.

Gu, C. (1989) RKPACK and its applications: fitting smoothing spline models. *Proc. Statist. Comput. Sect. Am. Statist. Ass.*, 42-51.

——(1990) Adaptive spline smoothing in non-Gaussian regression models. *J. Am. Statist. Ass.*, **85**, 801-807.

——(1992) Diagnostics for nonparametric regression models with additive terms. *J. Am. Statist. Ass.*, **87**, in the press.

Gu, C., Bates, D., Chen, Z. and Wahba, G. (1989) The computation of GCV functions through Householder tridiagonalization with application to the fitting of interaction spline models. *SIAM J. Matrix Anal. Applic.*, **10**, 457-480.

Gu, C. and Wahba, G. (1991a) Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method. *SIAM J. Sci. Statist. Comput.*, **12**, 383-398.

——(1991b) Discussion of "Multivariate adaptive regression splines" by J. Friedman. *Ann. Statist.*, **19**, 115-123.

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

Hutchinson, M., Kalma, J. and Johnson, M. (1984) Monthly estimates of windspeed and wind run for Australia. *J. Climat.*, **4**, 311-324.

Kimeldorf, G. and Wahba, G. (1971) Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.*, **33**, 82-95.

Mate, L. (1989) *Hilbert Space Methods in Science and Engineering*. New York: Hilger.

Meinguet, J. (1979) An intrinsic approach to multivariate spline interpolation at arbitrary points. In *Polynomial and Spline Approximation* (ed. B. N. Sahney), pp. 163-190. Dordrecht: Reidel.

Micchelli, C. (1986) Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constr. Approx.*, **2**, 11-22.

Nychka, D. (1988) Bayesian confidence intervals for smoothing splines. *J. Am. Statist. Ass.*, **83**, 1134-1143.

del Pino, G. (1991) Colloquium talk at the University of Wisconsin, Madison.

Ramsay, J. O. and Dalzell, C. J. (1991) Some tools for functional data analysis (with discussion). *J. R. Statist. Soc.* B, **53**, 539-572.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989) Design and analysis of computer experiments. *Statist. Sci.*, **4**, 409-435.

Stewart, G. W. (1987) Collinearity and least square regression. *Statist. Sci.*, **2**, 68-100.

Stone, C. (1985) Additive regression and other nonparametric models. *Ann. Statist.*, **15**, 885–941.
Utreras, F. (1988) On generalized cross-validation for multivariate smoothing spline functions. *SIAM J. Sci. Statist. Comput.*, **8**, 603–643.
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.
——— (1983) Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. R. Statist. Soc.* B, **45**, 133–150.
——— (1986) Partial and interaction splines for the semiparametric estimation of functions of several variables. In *Computer Science and Statistics: Proc. 18th Symp. Interface* (ed. T. J. Boardman), pp. 75–80. New York: American Statistical Association.
——— (1990) Spline models for observational data. *CBMS-NSF Reg. Conf. Ser. Appl. Math.*, **59**.
Wahba, G. and Wendelberger, J. (1980) Some new mathematical methods for variational objective analysis using splines and cross-validation. *Mthly Weath. Rev.*, **108**, 1122–1145.