SEMIPARAMETRIC ANOVA WITH TENSOR PRODUCT THIN PLATE SPLINES

by

Chong Gu and Grace Wahba
Purdue University and University of Wisconsin

Technical Report #90-61

Department of Statistics
Purdue University

November 1990

Semiparametric ANOVA with Tensor Product Thin Plate Splines

CHONG GU and GRACE WAHBA*

November 1990

Abstract

Thin plate splines and interaction splines are two different multidimensional spline smoothing techniques, the former being of radial structure and the latter being of tensor product structure. In this article we discuss a synthesis of the two methods, where the latter serves as a super structure under which the former acts as axes. 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.

KEY WORDS: Model selection; Multidimensional smoothing splines; Reproducing kernel; Spatial data analysis; Spline ANOVA decomposition.

---

*Chong Gu is Assistant Professor, Department of Statistics, Purdue University, West Lafayette, IN 47907. Grace Wahba is John Bascom Professor, Department of Statistics, University of Wisconsin, Madison, WI 53706. This research was supported by National Science Foundation under Grant DMS-9002566, Air Force Office of Scientific Research under Grant AFOSR-90-0103, and National Aeronautics and Space Administration under Contract NAG5-316

1 INTRODUCTION

Suppose we observe

$$y_j = f(t_j) + \epsilon_j, \qquad j = 1, \cdots, n$$

where $t_j \in R^d$ and the ϵ_j's are *i.i.d.* noise with mean 0 and variance σ^2. We want to estimate f from the data. The smoothing spline method estimates f using the minimizer of

$$\frac{1}{n}\sum_{j=1}^{n}(y_j - f(t_j))^2 + \lambda J(f) \tag{1.1}$$

subject to $f \in \mathcal{H}$, where the square error measures the goodness-of-fit and the $J(f)$ is a roughness penalty, the smoothing parameter λ controls the tradeoff of the two, and the $\mathcal{H}$ is a function space of tentative models. Usually the $J(f)$ and the $\mathcal{H}$ come as a pair.

Thin plate splines result from taking $\mathcal{H}$ to be all functions with square integrable derivatives of a fixed *total* degree m, say, and taking

$$J_m^d(f) = \sum_{\alpha_1+\cdots+\alpha_d=m} \frac{m!}{\alpha_1!\cdots\alpha_d!}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{\partial^m f}{\partial t_1^{\alpha_1}\cdots\partial t_d^{\alpha_d}}\right)^2 dt_1\cdots dt_d. \tag{1.2}$$

A technical requirement is $2m > d$. The penalty J_m^d is invariant under a rotation of variables, 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, or latitude-longitude-altitude for atmospheric data, or latitude-longitude-depth for oceanic data. As a matter of fact, the thin plate spline estimate of f with the penalty J_m^d is the sum of a lower degree polynomial (with the *total* degree less than m) and a linear combination of radial functions $E_m^d(\|\cdot - t_j\|)$, where E_m^d is a known univariate function (see next section) and $\|\cdot\|$ is the Euclidean norm in R^d.

The $\mathcal{H}$ for a "classical" interaction spline is a function space of functions which have a unique decomposition of the form $f(t) = C + \sum_\gamma f_\gamma(t_\gamma) + \sum_{\gamma_1<\gamma_2} f_{\gamma_1,\gamma_2}(t_{\gamma_1}, t_{\gamma_2}) + \cdots + f_{1,\cdots,d}(t_1, \cdots, t_d)$. The components are in orthogonal subspaces in $\mathcal{H}$ and orthogonality conditions impose side conditions which insure uniqueness, analogous to parametric ANOVA. Each function has a *marginal* square integrable m_γth derivative with respect to t_γ, where the m_γ's may be different. The f_γ's are the *main effects*, the f_{γ_1,γ_2}'s are the *two-factor interactions*, and so on. This decomposition is obtained by letting $\mathcal{H}$ be the tensor product of spaces of functions of one variable, and expanding this tensor product into tensor sums of subspaces according to the desired ANOVA decomposition. The $J(f)$ is a (weighted) sum of the roughness penalties on the different components of f. Such a modularity allows fairly flexible model specifications. For example, deleting higher order interactions leads to less adaptive but more "estimable" models, and the main-effect-only models reduce to the popular additive models. The tensor product structure is natural when all the d variables are qualitatively different, and the ANOVA decomposition greatly enhances the interpretability of the estimated functions. Interaction spline estimate of f can be written as linear combinations of products $\prod_{\gamma=1}^{d} g_\gamma(t_\gamma)$'s, where the g_γ's are either polynomials of degree less than m_γ or functions $E^1_{m_\gamma}(|\cdot - t_{\gamma,j}|)$.

In certain applications, the d variables may be partitioned into qualitatively homogeneous groups, say $\boldsymbol{t} = (\{\boldsymbol{t}^{(1)}\}^T, \{\boldsymbol{t}^{(2)}\}^T)^T$ for example, and a natural structure should be invariant under within-group variable rotations. To obtain the modularity and interpretability of an ANOVA decomposition of $f(\boldsymbol{t}^{(1)}, \boldsymbol{t}^{(2)})$, however, a tensor product structure between groups is desired. The purpose of this article is to spell out how this can be done. The functions we will be getting can be written as linear combinations of functions of form $g_1(\boldsymbol{t}^{(1)})g_2(\boldsymbol{t}^{(2)})$, where g_γ's are either polynomials of total degree less than m_γ or functions $E^{d_\gamma}_{m_\gamma}(\|\cdot - \boldsymbol{t}^{(\gamma)}_j\|)$, $\gamma = 1, 2$. Generalization will be immediate for more than two qualitatively homogeneous groups.

This work represents something of a synthesis and generalization of results in several branches of the statistical and approximation-theoretic literature. General theory of additive splines is to be found in Buja, Hastie, and Tibshirani (1989) and the book by Hastie and Tibshirani (1990). Related references to interaction splines include Barry (1986), Wahba (1986), Chen, Gu, and Wahba (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, Kalma, and Johnson (1984). Work related to data based choice of multiple smoothing parameters is found in Gu, Bates, Chen, and Wahba (1989), Gu (1989) and Gu and Wahba (1991 a). We use here the model diagnostics developed by Gu (1990 b). A brief preliminary sketch of the main idea here appears in Gu and Wahba (1991 b). A one-dimensional special case of the trick we use to define reproducing kernels below 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, MARS (Friedman 1991) and the $\prod$-method (Breiman 1991).

The rest of the article is organized as follows. In section 2, we describe the general method of constructing tensor product thin plate splines, where the reproducing kernels play a central role. In section 3, we describe a sample model in some detail. The calculation of the general model is sketched in Section 4. An analysis of an application which motivated this research is presented in Section 5. Finally, Section 6 concludes the article with discussions.

2 GENERAL FORMULATION AND REPRODUCING KERNELS

We now formalize (1.1) on an arbitrary domain $\mathcal{T}$. It will be assumed that $\mathcal{H}$ is a Hilbert space in which any evaluation functional $[t]f = f(t)$ is continuous. This is necessary since we are sampling f at arbitrary points and we wouldn't regard f and g to be close when $f(t)$ and $g(t)$ are far apart at a point t. Such a space is called a *reproducing kernel Hilbert space* possessing a *reproducing kernel* $R(\cdot,\cdot)$ such that $R(t,\cdot) \in \mathcal{H}$, $\forall t \in \mathcal{T}$, and $< R(t,\cdot), f >= f(t)$, $\forall f \in \mathcal{H}$, where $< \cdot,\cdot >$ is the inner product in $\mathcal{H}$. As a matter of fact, all members of the space $\mathcal{H}$ can be represented as linear combinations of $R(t,\cdot)$, $t \in \mathcal{H}$, or their limits. $J(f)$ is usually taken as a semi-norm in $\mathcal{H}$ having a (finite dimensional) null space $\mathcal{H}_0$. The boundary conditions for making $J(f)$ a norm and a class of equivalent norms on $\mathcal{H}_0$ determine each other uniquely, leading to a tensor sum decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$. The reproducing kernel can be decomposed accordingly as $R = R_0 + R_1$, where R_i are reproducing kernels of $\mathcal{H}_i$, $i = 0, 1$, respectively.

On a product domain $\mathcal{T} = \mathcal{T}^{(1)} \times \mathcal{T}^{(2)}$, given reproducing kernel Hilbert spaces $\mathcal{H}^{(1)}$ on $\mathcal{T}^{(1)}$ with reproducing kernel $R^{(1)}$ and $\mathcal{H}^{(2)}$ on $\mathcal{T}^{(2)}$ with reproducing kernel $R^{(2)}$, one can construct a tensor product reproducing kernel Hilbert space $\mathcal{H} = \mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$ with a reproducing kernel $R(\{t^{(1)}, t^{(2)}\}, \{s^{(1)}, s^{(2)}\}) = R^{(1)}(t^{(1)}, s^{(1)})R^{(2)}(t^{(2)}, s^{(2)})$, where $t^{(1)}, s^{(1)} \in \mathcal{T}^{(1)}$ and $t^{(2)}, s^{(2)} \in \mathcal{T}^{(2)}$. The decompositions of $\mathcal{H}^{(\gamma)} = \mathcal{H}_0^{(\gamma)} \oplus \mathcal{H}_1^{(\gamma)}$, $\gamma = 1, 2$, lead to a decomposition $\mathcal{H} = (\mathcal{H}_0^{(1)} \otimes \mathcal{H}_0^{(2)}) \oplus (\mathcal{H}_1^{(1)} \otimes \mathcal{H}_0^{(2)}) \oplus (\mathcal{H}_0^{(1)} \otimes \mathcal{H}_1^{(2)}) \oplus (\mathcal{H}_1^{(1)} \otimes \mathcal{H}_1^{(2)}) = \mathcal{H}_{0,0} \oplus \mathcal{H}_{1,0} \oplus \mathcal{H}_{0,1} \oplus \mathcal{H}_{1,1}$, with reproducing kernels $R_{\alpha,\beta} = R_\alpha^{(1)} R_\beta^{(2)}$, $\alpha, \beta = 0, 1$. A smoothing spline on $\mathcal{T}$ can then be defined by taking $J(f) = \theta_{1,0}^{-1}\|P_{1,0}f\|_{1,0}^2 + \theta_{0,1}^{-1}\|P_{0,1}f\|_{0,1}^2 + \theta_{1,1}^{-1}\|P_{1,1}f\|_{1,1}^2$, where $P_{\alpha,\beta}$ are projections onto $\mathcal{H}_{\alpha,\beta}$ and

$\|\cdot\|_{\alpha,\beta}$ are the norms on $\mathcal{H}_{\alpha,\beta}$ associated with $R_{\alpha,\beta}$. $J(f)$ has a finite dimensional null space $\mathcal{H}_{0,0}$ and is a norm on $\mathcal{H}_{1,0}\oplus\mathcal{H}_{0,1}\oplus\mathcal{H}_{1,1}$ with associated reproducing kernel $\theta_{1,0}R_{1,0}+\theta_{0,1}R_{0,1}+\theta_{1,1}R_{1,1}$. We remark that the developments presented above extend obviously to multiway tensor products and multiterm marginal tensor sums. See, e.g., Aronszajn (1950), Chen et al. (1989), Wahba (1986, 1990), and Gu et al. (1989).

With J_m^d as a semi-norm, it is known that an appropriate function space $\mathcal{X}$ (cf. Meinguet 1979) of functions with $J_m^d < \infty$ $(2m > d)$ can be made a reproducing kernel Hilbert space by defining a norm on the null space $\mathcal{H}_0$ of J_m^d. This null space is the $M = M(m,d) = \begin{pmatrix} m+d-1 \\ d \end{pmatrix}$ polynomials of total degree less than m in d variables. See also Wahba and Wendelberger (1980). Our remaining task is to define a useful norm on $\mathcal{H}_0$ for our purpose and to work out the associated reproducing kernels. Since polynomials are not integrable on R^d, one choice is to define the norm in terms of sums over a set of points $\{\boldsymbol{w}_k\}_{k=1}^N$, call it a normalizing mesh. Now let $(f,g)_N = (1/N)\sum_{k=1}^N f(\boldsymbol{w}_k)g(\boldsymbol{w}_k)$ and let $\{\phi_\nu\}_{\nu=0}^{M-1}$ be a set of bases of $\mathcal{H}_0$ satisfying $\phi_0 = 1$, $(\phi_\nu,\phi_\nu)_N = 1$, and $(\phi_\nu,\phi_\mu)_N = 0$, $\nu \neq \mu$. This is only possible when the restriction of $\mathcal{H}_0$ to the normalizing mesh keeps its dimension M. Defining the inner product on $\mathcal{H}_0$ as $< f,g >_0 = (f,g)_N$, $f,g \in \mathcal{H}_0$, it is easily seen that the projection of $f \in \mathcal{H}$ onto $\mathcal{H}_0$ is $(P_0 f)(\cdot) = \sum_{\nu=0}^{M-1}(\phi_\nu,f)_N\phi_\nu(\cdot)$. Partitioning $\{f:\ J_m^d(f) < \infty\} = \mathcal{H}_c\oplus\mathcal{H}_\pi\oplus\mathcal{H}_s$, where $\mathcal{H}_c = \{1\}$, $\mathcal{H}_\pi = \{\phi_1,\cdots,\phi_{M-1}\}$ (hence $\mathcal{H}_0 = \mathcal{H}_c\oplus\mathcal{H}_\pi$), and $\mathcal{H}_s = \{f:\ P_0 f = 0, J_m^d < \infty\}$, we have the following theorem.

Theorem 2.1 *Let I be the identity operator and $P_{0(\boldsymbol{t})}$ be P_0 acting on what follows considered as a function of $\boldsymbol{t}$. Associated with the square norm $\|f\|^2 = (P_0 f, P_0 f)_N + J_m^d(f)$, the reproducing kernels of $\mathcal{H}_c$, $\mathcal{H}_\pi$, and $\mathcal{H}_s$ are $R_c(\boldsymbol{t},\boldsymbol{s}) = 1$, $R_\pi(\boldsymbol{t},\boldsymbol{s}) = \sum_{\nu=1}^{M-1}\phi_\nu(\boldsymbol{t})\phi_\nu(\boldsymbol{s})$, and $R_s(\boldsymbol{t},\boldsymbol{s}) = (I - P_{0(\boldsymbol{t})})(I - P_{0(\boldsymbol{s})})E(\boldsymbol{t},\boldsymbol{s})$, respectively, where $E(\boldsymbol{t},\boldsymbol{s}) = E_m^d(\|\boldsymbol{t}-\boldsymbol{s}\|)$ and*

$$E_m^d(\cdot) = \begin{cases} C_m\{(\cdot)^{2m-d}\log(\cdot)\}, & d \text{ even}, \\ & C_m = (-1)^{d/2+m+1} \,/\, \{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!\} \\ C_m\{(\cdot)^{2m-d}\}, & d \text{ odd}, \\ & C_m = \Gamma(d/2-m) \,/\, \{2^{2m}\pi^{d/2}(m-1)!\} \end{cases}$$

The proof of the theorem is given in the appendix. Using these reproducing kernels to construct a tensor product reproducing kernel Hilbert space on $R^{d_1}\times\cdots\times R^{d_K}$ with marginal orders m_γ, $\gamma = 1,\cdots,K$, say, we get tensor product thin plate splines. Our construction here depends on the

Table 3.1: Terms in tensor product space $\mathcal{H}$.

$\mathcal{H}_{c,c}$	$\mathcal{H}_{c,\pi}$	$\mathcal{H}_{c,s}$
$\mathcal{H}_{\pi,c}$	$\mathcal{H}_{\pi,\pi}$	$\mathcal{H}_{\pi,s}$
$\mathcal{H}_{s,c}$	$\mathcal{H}_{s,\pi}$	$\mathcal{H}_{s,s}$

choice of the marginal normalizing meshes and results in a ready ANOVA decomposition whose side conditions are consistent with the normalizing meshes; see Section 3. Taking the marginal normalizing meshes as the marginals of the sampling points $\{t_j\}_{j=1}^n$ seems natural for many applications and also leads to convenient calculation; see Section 4. We remark that the special case of $d = 1$ can be shown to lead to the usual polynomial smoothing splines of degree $2m - 1$.

3 A SAMPLE MODEL

We describe in detail the construction of a sample model for $f(t_1, t_2, t_3) = f(t^{(1)}, t^{(2)})$ on $\mathcal{T}^{(1)} \times \mathcal{T}^{(2)} = R \times R^2$, where we take $m_1 = m_2 = 2$. On R, we have $\mathcal{H}^{(1)} = \mathcal{H}_c^{(1)} \oplus \mathcal{H}_\pi^{(1)} \oplus \mathcal{H}_s^{(1)}$, where $\mathcal{H}_c^{(1)} = \{1\}$ contains constants, $\mathcal{H}_\pi^{(1)} = \{\phi^{(1)}\}$ contains linear functions summing to zero over the marginal normalizing mesh $\{w_k^{(1)}\}_{k=1}^{N_1}$, and $\mathcal{H}_s^{(1)} = \{f : \sum_k f(w_k^{(1)}) = \sum_k \phi^{(1)}(w_k^{(1)}) f(w_k^{(1)}) = 0, \int_{-\infty}^{\infty} \ddot{f}^2 < \infty\}$ (the "smooth" part). On R^2, we have $\mathcal{H}^{(2)} = \mathcal{H}_c^{(2)} \oplus \mathcal{H}_\pi^{(2)} \oplus \mathcal{H}_s^{(2)}$, where $\mathcal{H}_c^{(2)} = \{1\}$ contains constants, $\mathcal{H}_\pi^{(2)} = \{\phi_1^{(2)}, \phi_2^{(2)}\}$ contains linear polynomials summing to zero over the marginal normalizing mesh $\{w_k^{(2)}\}_{k=1}^{N_2}$, and $\mathcal{H}_s^{(2)} = \{f : \sum_k f(w_k^{(2)}) = \sum_k \phi_1^{(2)}(w_k^{(2)}) f(w_k^{(2)}) = \sum_k \phi_2^{(2)}(w_k^{(2)}) f(w_k^{(2)}) = 0, \int\int(\{\partial^2 f/\partial t_2^2\}^2 + 2\{\partial^2 f/\partial t_2 \partial t_3\}^2 + \{\partial^2 f/\partial t_3^2\}^2) dt_2 dt_3 < \infty\}$. The tensor product space $\mathcal{H} = \mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$ has accordingly a tensor sum decomposition of nine terms as in Table 3.1. where $\mathcal{H}_{\alpha,\beta} = \mathcal{H}_\alpha^{(1)} \otimes \mathcal{H}_\beta^{(2)}$, $\alpha, \beta = c, \pi, s$. The four terms on the upper-left corner have finite dimensions, and by convention are not penalized (i.e., the corresponding $\theta = \infty$) in most applications. The other five terms all have infinite dimension and have to be penalized in a fit.

Defining the side conditions according to the normalizing meshes, i.e., that both main effects sum to zero on the corresponding marginal normalizing meshes and the interaction sums to zero over the marginal normalizing mesh on both axes, the ANOVA decomposition is readily available from the decomposition of $\mathcal{H}$ in Table 3.1. Specifically, f_1 comes from $\mathcal{H}_{\pi,c} \oplus \mathcal{H}_{s,c}$, f_2 comes from $\mathcal{H}_{c,\pi} \oplus \mathcal{H}_{c,s}$, and $f_{1,2}$ comes from the four terms on the lower-right corner of Table 3.1. By eliminating these four terms (i.e., setting the corresponding $\theta = 0$) one obtains an additive model.

We note that the grouping of the nine terms in model fitting and model interpretation are different.

Finally let us look at the penalties. It can be shown that the norm on $\mathcal{H}_{s,c}$ is proportional to $\int \left(\{\partial^2/\partial t_1^2\}(1/N_2)\sum_k f(t_1, w_k^{(2)})\right)^2 dt_1$. Noticing that the $t^{(1)}$ main effect is $f_1(t_1) = (1/N_2)\sum_k f(t_1, w_k^{(2)}) - (1/N_1N_2)\sum_{j,k} f(w_j^{(1)}, w_k^{(2)})$, we are reassured that the penalty on $\mathcal{H}_{s,c}$ is the usual marginal penalty acting on the main effect. Similarly, the norm on $\mathcal{H}_{c,s}$ is proportional to

$$\int\int \left(\frac{\partial^2 f_2}{\partial t_2^2}\right)^2 + 2\left(\frac{\partial^2 f_2}{\partial t_2 \partial t_3}\right)^2 + \left(\frac{\partial^2 f_2}{\partial t_3^2}\right)^2 dt_2 dt_3,$$

where $f_2(t_2, t_3) = (1/N_1)\sum_k f(w_k^{(1)}, t_2, t_3) - (1/N_1N_2)\sum_{j,k} f(w_j^{(1)}, w_k^{(2)})$ is the $t^{(2)}$ main effect. In general there are three penalty terms for the interaction, acting on polynomial-smooth $(\mathcal{H}_{\pi,s})$, smooth-polynomial $(\mathcal{H}_{s,\pi})$, and smooth-smooth $(\mathcal{H}_{s,s})$ interactions. The norm on $\mathcal{H}_{s,\pi}$ is proportional to

$$\int \left(\frac{\partial^2}{\partial t_1^2}\sum_k \{\phi_1^{(2)}(w_k^{(2)}) + \phi_2^{(2)}(w_k^{(2)})\} f(t_1, w_k^{(2)})\right)^2 dt_1,$$

the norm on $\mathcal{H}_{\pi,s}$ is similar, and the norm on $\mathcal{H}_{s,s}$ is proportional to

$$\int\int \left(\frac{\partial^4 f}{\partial t_1^2 \partial t_2^2}\right)^2 + 2\left(\frac{\partial^4 f}{\partial t_1^2 \partial t_2 \partial t_3}\right)^2 + \left(\frac{\partial^4 f}{\partial t_1^2 \partial t_3^2}\right)^2 dt_2 dt_3.$$

4 CALCULATION

Tensor product thin plate splines are special cases of smoothing splines with multiple smoothing parameters, and as such, their computation can be conducted using the generic algorithms of Gu and Wahba (1991 a). The reproducing kernels of Theorem2.1, however, depend on the choice of the normalizing mesh, hence some preliminary calculation is necessary before applying the algorithms.

Without loss of generality we consider tensor products of two marginals. It can be shown that the minimizer of (1.1) with $J(f) = \sum_{s\in\{\alpha,\beta\}} \theta_{\alpha,\beta}^{-1} \|P_{\alpha,\beta} f\|_{\alpha,\beta}$ has an expression

$$f(\cdot) = \sum_{\nu=0}^{M-1} d_\nu \phi_\nu(\cdot) + \sum_{j=1}^{n} c_j\Big(\sum_{s\in\{\alpha,\beta\}} \theta_{\alpha,\beta} R_{\alpha,\beta}(t_j, \cdot)\Big). \tag{4.1}$$

Here $M = M(m_1, d_1)M(m_2, d_2)$, each ϕ_ν is of the form $\phi_\nu(t) = \phi_{\nu_1}(t^{(1)})\phi_{\nu_2}(t^{(2)})$, $\alpha, \beta = c, \pi, s$, $\{\alpha, \beta\}$ indexes the subspaces which are included and penalized, and $R_{\alpha,\beta}(t, s) = R_\alpha^{(1)}(t^{(1)}, s^{(1)})$

$R_\beta^{(2)}(\boldsymbol{t}^{(2)}, \boldsymbol{s}^{(2)})$. $\boldsymbol{c}$ and $\boldsymbol{d}$ are solutions to the linear system

$$\begin{aligned} (\{\sum_{s\in\{\alpha,\beta\}} \theta_{\alpha,\beta} Q_{\alpha,\beta}\} + n\lambda I)\boldsymbol{c} + S\boldsymbol{d} &= \boldsymbol{y} \\ S^T\boldsymbol{c} &= 0, \end{aligned} \tag{4.2}$$

where $Q_{\alpha,\beta}$ and S are $n \times n$ and $n \times M$ matrices, defined by $Q_{\alpha,\beta} = (R_{\alpha,\beta}(\boldsymbol{t}_i, \boldsymbol{t}_j))$ and $S = (\phi_\nu(\boldsymbol{t}_j))$. It can be seen that $Q_{\alpha,\beta} = Q_\alpha^{(1)} \circ Q_\beta^{(2)}$, where $\circ$ indicates component-wise multiplication and $Q_\alpha^{(1)} = R_\alpha^{(1)}(t_i^{(1)}, t_j^{(1)})$ and $Q_\beta^{(2)} = R_\beta^{(2)}(t_i^{(2)}, t_j^{(2)})$. Similarly, columns of S are the component-wise products of all column combinations of matrices $S^{(1)}$ and $S^{(2)}$, where $S^{(1)} = (\phi_\nu^{(1)}(t_j^{(1)}))$ and $S^{(2)} = (\phi_\nu^{(2)}(t_j^{(2)}))$. The algorithms of Gu and Wahba (1991 a) are designed for solving (4.2) with the $\lambda/\theta_{\alpha,\beta}$'s selected using the generalized cross validation criterion of Craven and Wahba (1979), given S, $Q_{\alpha,\beta}$'s, and $\boldsymbol{y}$, so the preliminary calculation we need is to generate the marginal matrices $Q_\alpha^{(1)}$, $Q_\beta^{(2)}$, $S^{(1)}$, and $S^{(2)}$.

In the rest of the section, we let Q_α and S stand for $Q_\alpha^{(\gamma)}$ and $S^{(\gamma)}$, $\gamma = 1, 2$, and we drop superscripts and boldface notations for points $\boldsymbol{t}^{(\gamma)}$ and $\boldsymbol{w}^{(\gamma)}$ on the (marginal) domain. It can be seen that $Q_c = \mathbf{1}\mathbf{1}^T$, $Q_\pi = S_tS_t^T - Q_c$, and $Q_s = K_{t,t} - (S_tS_w^T/N)K_{w,t} - K_{t,w}(S_wS_t^T/N) + (S_tS_w^T/N)K_{w,w}(S_wS_t^T/N)$, where $K_{u,v} = (E_m(u_i, v_j))$, $S_u = (\phi_\nu(u_j))$, u and v are to be replaced by t and w, and t_j's are the design points and w_k's are the normalizing mesh. To determine the ϕ_ν's and to calculate S_w and S_t, we take a convenient basis $\{\psi_\nu\}$ for the null space with $\psi_0 = \phi_0 = 1$ and formulate matrices $\tilde{S}_w$ and $\tilde{S}_t$ by replacing the ϕ_ν with ψ_ν in the definition of S_u, then from the QR-decomposition (without pivoting) $\tilde{S}_w = (F_1, F_2)\begin{pmatrix} R_1 \\ 0 \end{pmatrix} = F_1R_1$ we get $S_w/\sqrt{N} = F_1$ and $S/\sqrt{N} = S_t/\sqrt{N} = \tilde{S}_tR_1^{-1}$, where $\boldsymbol{\phi} = \sqrt{N}R_1^{-T}\boldsymbol{\psi}$. For evaluating the model at an arbitrary point u, $K_{t,u} - (S_tS_w^T/N)K_{w,u} - K_{t,w}(S_wS_u^T/N) + (S_tS_w^T/N)K_{w,w}(S_wS_u^T/N)$ is needed, and the matrices $S_t/\sqrt{N}$, $S_w/\sqrt{N}$, $K_{t,w}$, and $K_{w,w}$ need to be maintained.

When $\{w_k\} = \{t_j\}$, $Q_s = (I - S_tS_t^T/N)K_{t,t}(I - S_tS_t^T/N) = F_2F_2^TK_{t,t}F_2F_2^T$, the calculation is significantly simplified. Similarly, $K_{t,u} - (S_tS_t^T/N)K_{t,u} - K_{t,t}(S_tS_u^T/N) + (S_tS_t^T/N)K_{t,t}(S_tS_u^T/N) = (I - S_tS_t^T/N)(K_{t,u} - K_{t,t}\{S_tS_u^T/N\}) = F_2F_2^T(K_{t,u} - K_{t,t}\tilde{S}_tR_1^{-1}R_1^{-T}\tilde{S}_u)$, where $K_{t,t}\tilde{S}_tR_1^{-1}R_1^{-T}$ can be calculated once for all. We note that the F_2 need not to be explicitly formed. These formulas are especially suitable for efficient calculation using LINPACK facilities.

5 AN EXAMPLE

In this section, we analyze some environmental data which motivated our conception of tensor product thin plate splines. The data we will be analyzing are derived by Douglas and Delampady (1990) from the Eastern Lake Survey of 1984 implemented by the EPA. The derived data set contains geographic information, water acidity measurements, and main ion concentrations of 1798 lakes in three regions, Northeast, Upper Midwest, and Southeast, in the Eastern United States. 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 suggest 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 Southeast, 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 3 with the normalizing mesh given by the design points. t_1 was taken as the logarithm of the calcium concentration (mg/L). (t_2, t_3) were obtained by converting the longitude and the latitude of the lake location to the east-west and the north-south distances from a local center. The calculations of the models were performed as described in Section 4, making use of the generic algorithms of Gu and Wahba (1991 a) which are implemented in RKPACK (Gu 1989). With the automatic smoothing parameter selection by the generalized cross validation method of Craven and Wahba (1979), the fits are invariant to the scalings of the axis domains. The fitted models were then decomposed into a constant, two main effects, and an interaction, as described in Section 4.

Evaluating a computed model at the design points, we get a retrospective linear model $\boldsymbol{y} = \tilde{\boldsymbol{f}}_{0,0} + \tilde{\boldsymbol{f}}_{1,0} + \tilde{\boldsymbol{f}}_{0,1} + \tilde{\boldsymbol{f}}_{1,1} + \tilde{e}$, and adjusting for the constant effect by projecting on to $\{\mathbf{1}\}^{\perp}$ we get $\boldsymbol{z} = \boldsymbol{f}_{1,0} + \boldsymbol{f}_{0,1} + \boldsymbol{f}_{1,1} + \boldsymbol{e}$. To measure the concurvity in the fit, we use the collinearity indices of Stewart (1987), $\kappa_i = \|\boldsymbol{f}_i\|\|\boldsymbol{f}_i^{(+)}\|$ where $f_i^{(+)}$ is the ith row of the Moore-Penrose inverse of $(\boldsymbol{f}_{1,0}, \boldsymbol{f}_{0,1}, \boldsymbol{f}_{1,1})$, which can be computed from $\cos(\boldsymbol{f}_i, \boldsymbol{f}_j)$. The $\boldsymbol{f}$'s are supposed to predict the "response" $\boldsymbol{z}$ so a near orthogonal angle between a $\boldsymbol{f}_i$ and $\boldsymbol{z}$ indicates a noise term. Signal terms should be reasonably orthogonal to the residuals hence a large cosine between a $\boldsymbol{f}_i$ and $\boldsymbol{e}$ makes a term suspect. $\cos(\boldsymbol{z}, \boldsymbol{e})$ and $R^2 = \|\boldsymbol{z} - \boldsymbol{e}\|^2/\|\boldsymbol{z}\|^2$ are informative *ad hoc* measures for the signal to noise ratio in the data. A *very* small norm of a $\boldsymbol{f}_i$ relative to that of $\boldsymbol{z}$ also indicates a negligible

Table 5.1: Diagnostics for Florida Model.

	$f_{1,0}$	$f_{0,1}$	$f_{1,1}$	e	z
κ.	1.07	1.13	1.11	$R^2 = 0.793$	
$\cos(z,\cdot)$	0.861	0.045	0.076	0.457	1
$\cos(e,\cdot)$	0.007	0.106	0.129	1	0.457
$\|\cdot\|$	14.53	2.62	2.23	6.53	15.77

Table 5.2: Diagnostics for Blue Ridge Model.

	$f_{1,0}$	$f_{0,1}$	$f_{1,1}$	e	z
κ.	1.08	1.07	1.03	$R^2 = 0.632$	
$\cos(z,\cdot)$	0.648	0.574	0.358	0.617	1
$\cos(e,\cdot)$	0.000	0.124	0.249	1	0.617
$\|\cdot\|$	2.45	1.44	1.14	2.07	4.10

term. More discussion of these diagnostics can be found in Gu (1990 b).

For the Florida model, the diagnostics are summarized in Table 5.1. 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 Figure 5.1, where a scatter plot of the data is superimposed. We concluded that the water acidity of the surveyed lakes in Florida didn't illustrate any spatial pattern other than uniformity.

For the Blue Ridge model the diagnostics are summarized in Table 5.2. All three components seem nonnegligible. The calcium main effect is plotted in Figure 5.2. The geography main effect is plotted in Figure 5.3. We plotted contours only where there are data. We have done this freehand, but we are developing a method which should give an objective measure of how far one can reasonably extends the model beyond the data. The crest of the Southern Blue Ridge mountains runs roughly from southwest to northeast with the highest peak slightly to the left of the center of the figure. 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 5.3. Again we would preserve all three terms but the diagnostics suggested that elevation was a bit short of replacing geography in modeling pH. We finally plotted the residuals from the calcium-geography model versus the residuals from the calcium-elevation model in Figure 5.4, to illustrate the spread of the residuals and to double check that the calcium-geography model was indeed a "refinement"

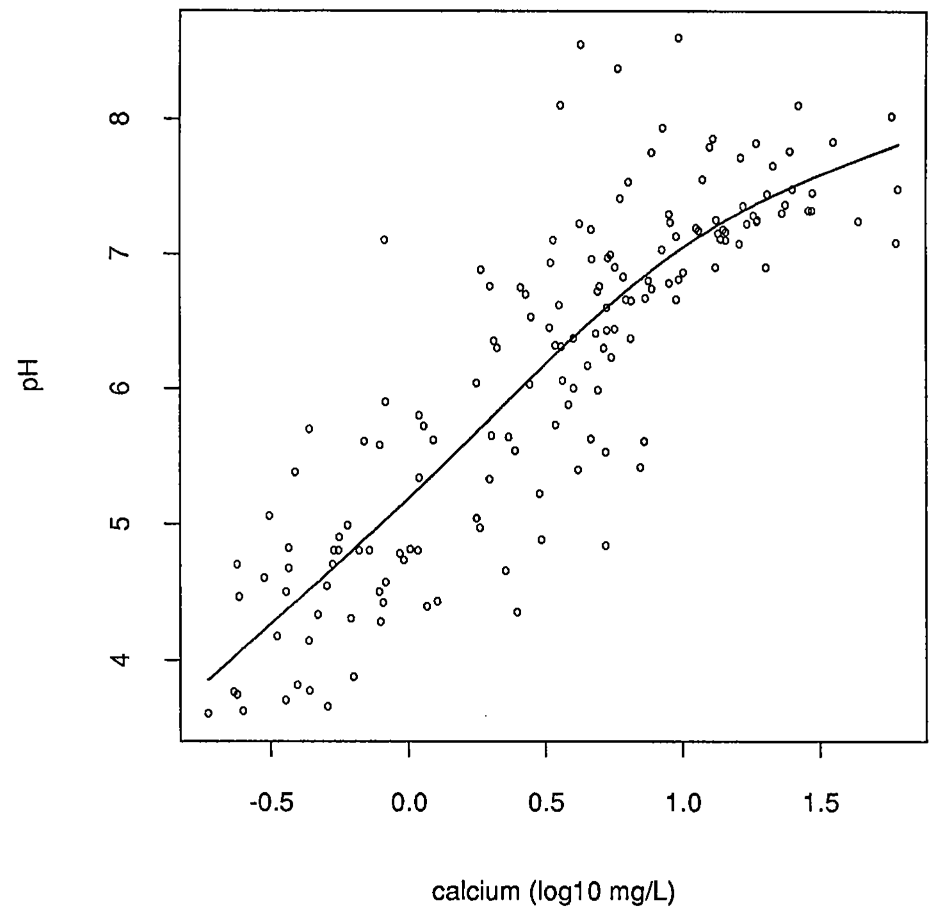

Figure 5.1: Final Florida Model.

Table 5.3: Diagnostics for Calcium-Elevation Model.

	$f_{1,0}$	$f_{0,1}$	$f_{1,1}$	e	z
$\kappa.$	1.40	1.17	1.22	$R^2 = 0.493$	
$\cos(z, \cdot)$	0.714	0.371	0.375	0.713	1
$\cos(e, \cdot)$	0.038	0.000	0.075	1	0.713
$\|\| \cdot \|\|$	2.50	0.59	0.26	2.81	4.10

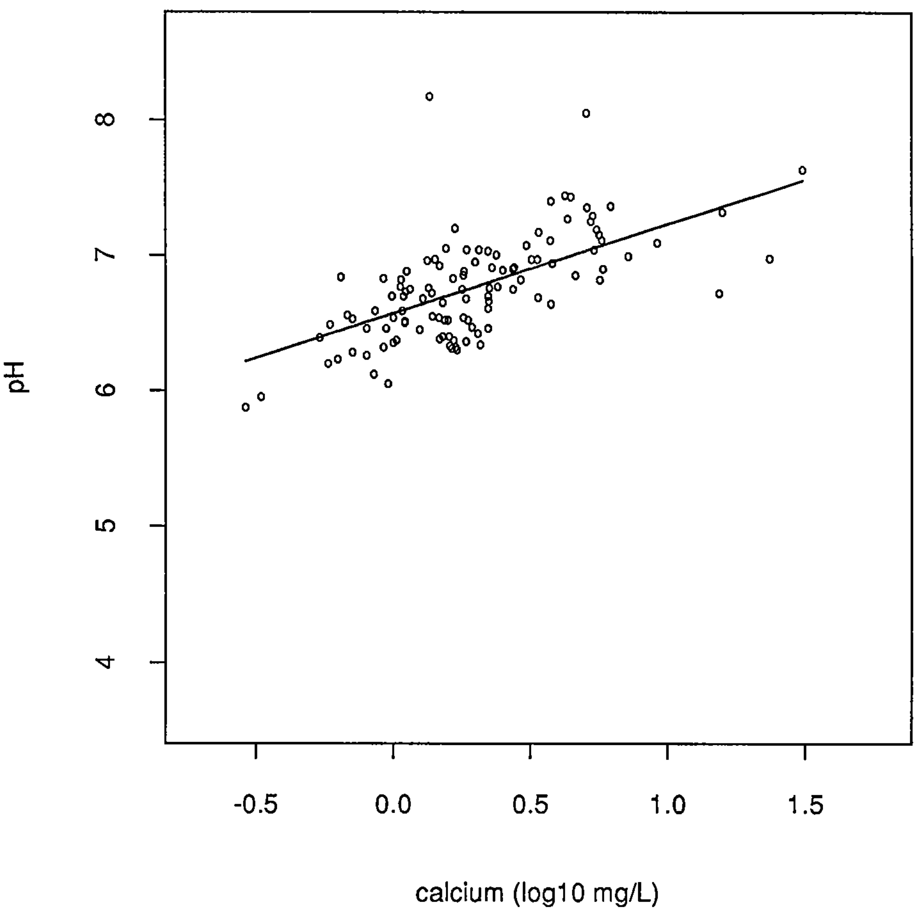

Figure 5.2: Calcium Main Effect of Blue Ridge Model.

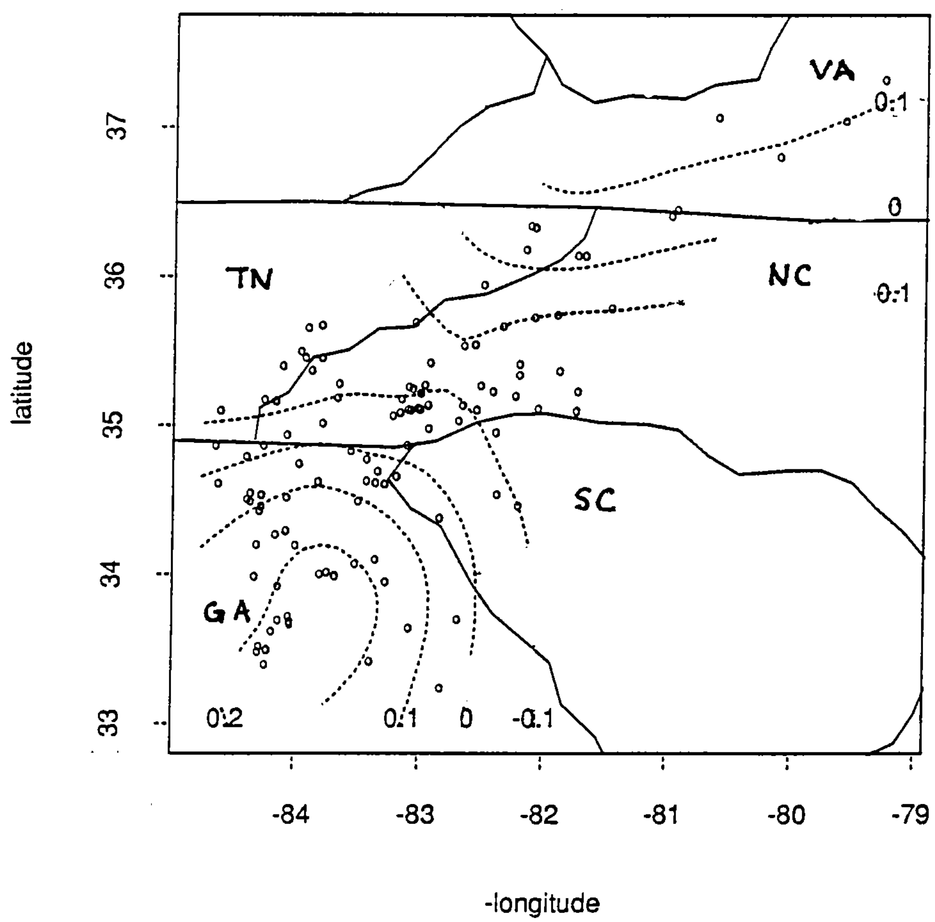

Figure 5.3: Geography Main Effect of Blue Ridge Model.

of the calcium-elevation model.

6 DISCUSSION

The models we construct in this article provide convenient nonparametric 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 one wishes to include, say a watershed factor, one simply includes 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 (that is, x is an indicator function for two watersheds). In dealing with three geographic variables, for example, longitude, latitude, and ocean depth, one 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 the generalized cross validation, see Hutchinson et al. (1984). Similar remarks may apply to time. Replacing mean square goodness-of-fit by mean minus log likelihood, the same structure can be used to fit odds ratios for binary data and to fit intensities for Poisson data, etc.; see, e.g., Gu (1990 a). Other extensions include models with inequality constraints, models based on aggregated data, etc; see Wahba (1990). For further technical remarks, see Appendix B.

Compared with model construction, model selection is much less understood. We have chosen to use the model selection procedures based on a retrospective linear model given in Gu (1990 b). Although the theory does not exist for determining when a $\cos(\boldsymbol{z}, \boldsymbol{f}_i)$ 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 generalized cross validation 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 generalized cross validation is a predictive criterion, and there appears to be a moderately large chance that the estimated θ_β is not zero even when the true f has $||P_\beta f|| = 0$ (Wahba 1990). Thus small components generated by noise may be retained. One could, in principle, develop an hypothesis testing approach for deciding when θ_β is *signifi-*

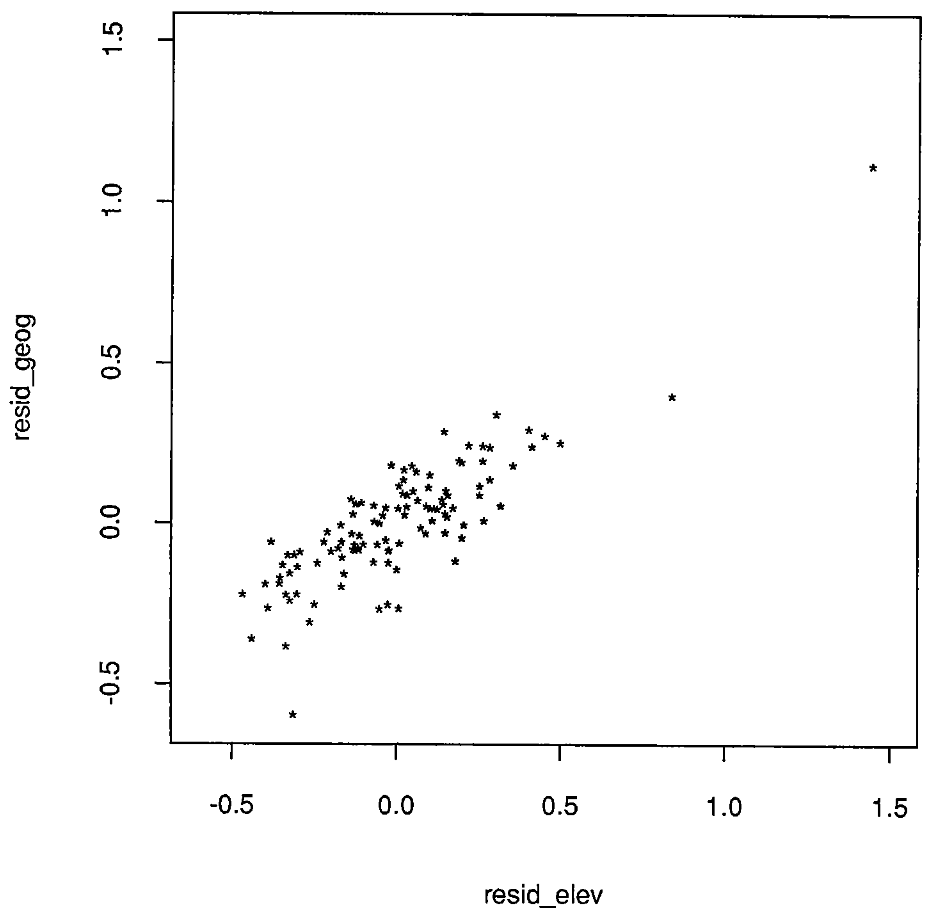

Figure 5.4: Residuals of Blue Ridge Models.

cantly non-zero, by looking at the statistic $v = V(\cdots;\infty)/V(\cdots;\lambda\hat{\theta}_\beta^{-1})$. Similarly, a likelihood ratio statistic $m = M(\cdots;\infty)/M(\cdots;\lambda\hat{\theta}_\beta^{-1})$ (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 θ's. Tests of the null hypothesis that the true model is in $\mathcal{H}_0$ are discussed in Cox, Koh, Wahba, and Yandell (1988) and Wahba (1990). Distributions under this (simple) null hypothesis are relatively straightforward to obtain by Monte Carlo methods.

A PROOF OF THEOREM 2.1

To show that $R(\boldsymbol{t},\boldsymbol{s})$ is the reproducing kernel for a given space, we need to show that $R(\boldsymbol{t},\cdot)$ is in the space, for each fixed $\boldsymbol{t}$, and that $< R(\boldsymbol{t},\cdot), R(\boldsymbol{s},\cdot) > = R(\boldsymbol{t},\boldsymbol{s})$. The assertions concerning R_c and R_π are immediate. The assertion concerning R_s will follow from a result of Meinguet (1979), which we give below after a few definitions. Given m and d, a set of $N+1$ points in R^d is said to be unisolvent, if least squares regression on $\phi_0,\cdots,\phi_{M-1}$ is unique. A set of (unisolvent) points $\boldsymbol{w}_0,\cdots,\boldsymbol{w}_N$ and associated weights $h_0,\cdots,h_N$ is called a generalized divided difference if

$$\sum_{k=0}^{N} h_k\phi_\nu(\boldsymbol{w}_k) = 0, \quad \nu = 0,1,\cdots,M-1.$$

Let

$$< f,g >_* = \sum_{\alpha_1+\cdots+\alpha_d=m} \frac{m!}{\alpha_1!\cdots\alpha_d!}\int\cdots\int \frac{\partial^m f}{\partial t_1^{\alpha_1}\cdots\partial t_d^{\alpha_d}}\frac{\partial^m g}{\partial t_1^{\alpha_1}\cdots\partial t_d^{\alpha_d}}dt_1\cdots dt_d.$$

We have the following Theorem.

Theorem A.1 (Meinguet 1979) *Let $\{h_k,\boldsymbol{w}_k\}$ be a generalized divided difference, and let $E(\boldsymbol{t},\boldsymbol{s})$ be defined as in the text. Let $E_{\boldsymbol{t}}(\cdot) = E(\boldsymbol{t},\cdot)$. Then*

$$< \sum_{j=0}^{N} h_j E_{\boldsymbol{w}_j}(\cdot), \sum_{k=0}^{N} h_k E_{\boldsymbol{w}_k}(\cdot) >_* = \sum_j\sum_k h_j h_k E(\boldsymbol{w}_j,\boldsymbol{w}_k).$$

We now proceed to the proof that R_s is the reproducing kernel for $\mathcal{H}_s$. Recalling that

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

we have to show that $R_s(t,\cdot)$ is perpendicular to $\mathcal{H}_0$ for each t, and

$$< R_s(t,\cdot), R_s(s,\cdot) >_* = R_s(t,s).$$

Since $P_{0(s)}$ is idempotent, it is obvious by construction that $P_{0(s)}R_s(t,s)$ is 0 for all t. We have that

$$R_s(t,\cdot) = E_t(\cdot) - \sum_{\nu=0}^{M-1} \phi_\nu(t) \sum_{k=1}^{N} \phi_\nu(w_k) E_{w_k}(\cdot) + \pi(\cdot),$$

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

$$E_t - \sum_{k=1}^{N} h_k(t) E_{w_k}(\cdot) + \pi(\cdot).$$

where

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

Now, we show that $(1, -h_1(t), \cdots, -h_N(t); t, w_1, \cdots, w_N)$ is a generalized divided difference. To do this we have to show that

$$\phi_\mu(t) - \sum_{k=1}^{N} h_k(t)\phi_\mu(t) = 0, \quad \mu = 0, \cdots, M-1.$$

Substituting in the definition of h_k the above becomes

$$\phi_\mu(t) - \sum_{\nu=0}^{M-1} \phi_\nu(t)(\phi_\nu, \phi_\mu)_N = 0.$$

which follows by the orthonormality of the $\{\phi_\nu\}$ under $(\cdot,\cdot)_N$. It then follows from Meinguet's theorem that

$$< R_s(t,\cdot), R_s(s,\cdot) >_* = < E_t(\cdot) - \sum_{\nu=0}^{M-1} \phi_\nu(t) \sum_{k=1}^{N} \phi_\nu(w_k) E_{w_k}(\cdot), E_s(\cdot) - \sum_{\nu=0}^{M-1} \phi_\nu(s) \sum_{k=1}^{N} \phi_\nu(w_k) E_{w_k}(\cdot) >_*,$$

$$= E(t,s) - \sum_{k=1}^{N} h_k(t) E(w_k, s) - \sum_{k=1}^{N} h_k(s) E(t, w_k) + \sum_{j=1}^{N}\sum_{k=1}^{N} h_j(t) h_k(s) E(w_j, w_k) = R_s(t,s),$$

so R_s is the reproducing kernel for $\mathcal{H}_s$ with the inner product $< \cdot,\cdot >_*$.

B TECHNICAL COMMENTS

We will not belabor the reader here with the relationship between splines and kriging. See the Letters to the Editor from Wahba and Cressie in The American Statistician, August, 1990 for different points of view, and references cited there for entree to the kriging literature. $E_m^d(\| \boldsymbol{t} - \boldsymbol{s} \|)$ is one of the so-called variograms in the kriging literature. In the definition of E_m^d, m does not have to be an integer. Replacing m by $m' = m + s$, the definition of $E_{m'}^d$ in Theorem 2.1 remains the same except that "d even" and "d odd" are replaced by "$2m' - d$ is an even integer" and "$2m' - d$ a positive real number not necessarily an integer". The norm in $\mathcal{H}_s$ for the thin plate spline with non-integer m' is defined in terms of Fourier transforms, see Duchon (1979) and Thomas-Agnan (1989). Of course when talking about polynomials, m must be an integer. Our construction here involving polynomials of total degree less than m on R^d will work if E_m^d is replaced by *any* function $F(\cdot)$ which is m-conditionally positive definite, see Micchelli (1986). F is m-conditionally positive definite if $\sum_{j,k} h_j h_k F(\|\boldsymbol{w}_j - \boldsymbol{w}_k\|) \geq 0$ whenever $\{h_k; \boldsymbol{w}_k\}$ is a generalized divided difference with respect to the polynomials of total degree less than m. This condition is exactly what one needs to show that $(I - P_{0(\boldsymbol{t})})(I - P_{0(\boldsymbol{s})})F(\boldsymbol{t}, \boldsymbol{s})$ is a non-negative definite function, i.e., a reproducing kernel. $E_{m'}^d$ is, according to Matheron (1973), m conditionally positive definite, for $0 \leq s < d/2$; see also Duchon (1977). It is $m+1$ conditionally positive definite whenever it is m conditionally positive definite, since this latter condition is stronger. Thus we could decouple the m in the definition of $\mathcal{H}_0$ and in the definition of $\mathcal{H}_s$ (subject to some restrictions), but we have chosen not to do so. We could in fact use $F(\|\boldsymbol{t} - \boldsymbol{s}\|) = \exp\{-\theta\|\boldsymbol{t} - \boldsymbol{s}\|^2\}$, which has been used by Sacks, Welch, Mitchell, and Wynn (1989) and others, or any other positive definite function, with any m. The polynomials can be replaced by other functions under certain conditions, see Kimeldorf and Wahba (1971), Wahba (1978), and Dalzell and Ramsay (1990). See Wahba (1990), Chapter 3, for a discussion on limiting the class of models one would want to consider.

References

Aronszajn, N. (1950), "Theory of Reproducing Kernels," *Transaction of the American Mathematical Society*, 68, 337 – 404.

Barry, D. (1986), "Nonparametric Bayesian Regression," *The Annals of Statistics*, 14, 934 – 953.

deBoor, C. and Lynch, R.E. (1966), "On Splines and Their Minimum Properties," *Journal of Mathematics and Mechanics*, 15, 953–969.

Breiman, L. (1991), "The $\prod$-Method for Estimating Multivariate Functions from Noisy Data," *Technometrics.*

Buja, A., Hastie, T., and Tibshirani, R. (1989), "Linear Smoothers and Additive Models" (with discussions), *The Annals of Statistics*, 17, 453 – 555.

Chen, Z., Gu, C., and Wahba, G. (1989), Discussion of "Linear Smoothers and Additive Models" by Buja, Hastie and Tibshirani, *The Annals of Statistics*, 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," *The Annals of Statistics*, 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," *Numerische Mathematik*, 31, 377 – 403.

Dalzell, C. J. and Ramsay, J. O. (1990), "Choosing an Inner Product for the Analysis of Functional Data," manuscript.

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, University of British Columbia, Dept. of Statistics.

Duchon, J. (1977), "Splines Minimizing Rotation-Invariant Semi-Norms in Sobolev Spaces," in *Constructive Theory of Functions of Several Variables*, pp. 85–100. Springer-Verlag.

Friedman, J. (1991), "Multivariate Adaptive Regression Splines," *The Annals of Statistics.*

Friedman, J. and Silverman, B. (1989), "Flexible Parsimonious Smoothing and Additive Modeling," *Technometrics*, 3, 3–21.

Gu, C. (1989), "RKPACK and Its Applications: Fitting Smoothing Spline Models," *Proceedings of Statistical Computing Section: American Statistical Association*, 42 – 51.

—— (1990 a), "Adaptive Spline Smoothing in Non-Gaussian Regression Models," *Journal of the American Statistical Association*, 85, 801 – 807.

—— (1990 b), "Diagnostics for Nonparametric Additive Models," Technical Report 92, University of British Columbia, Dept. of Statistics.

Gu, C., Bates, D. M., Chen, Z., and Wahba, G. (1989), "The Computation of GCV Functions through Householder Tridiagonalization with Application to the Fitting of Interaction Spline Models," *SIAM Journal on Matrix Analysis and Applications*, 10, 457 – 480.

Gu, C. and Wahba, G. (1991 a), "Minimizing GCV/GML Scores with Multiple Smoothing Parameters via the Newton Method," *SIAM Journal on Scientific and Statistical Computing.*

—— (1991 b), Discussion of "Multivariate Adaptive Regression Splines" by J. Friedman, *The Annals of Statistics.*

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

Hutchinson, M., Kalma, J., and Johnson, M. (1984), "Monthly Estimates of Windspeed and Wind Run for Australia," *Journal of Climatology*, 4, 311 – 324.

Kimeldorf, G. and Wahba, G. (1971), "Some Results on Tchebycheffian Spline Functions," *Journal of Mathematical Analysis and Applications*, 33, 82–95.

Matheron, G. (1973), "The Intrinsic Random Functions and Their Applications," *Advances in Applied Probability*, 5, 439–468.

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. Reidel.

Micchelli, C. (1986), "Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions," *Constructive Approximation*, 2, 11 – 22.

Nychka, D. (1988), "Bayesian Confidence Intervals for Smoothing Splines," *Journal of the American Statistical Association*, 83, 1134 – 1143.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments," *Statistical Science*, 4, 409–435.

Stewart, G. W. (1987), "Collinearity and Least Square Regression," *Statistical Science*, 2, 68 – 100.

Thomas-Agnan, C. (1989), "Smoothing Noisy Data by Two Equivalent Techniques," in *Data Analysis, Learning Symbolic and Numeric Knowledge*, ed. E. Diday, pp. 000-000. Nova Science Publishers.

Utreras, F. (1988), "On Generalized Cross-Validation for Multivariate Smoothing Spline Functions," *SIAM Journal on Scientific and Statistical Computing*, 8, 630–643.

Wahba, G. (1978), "Improper Priors, Spline Smoothing and the Problem of Guarding against Model Errors in Regression," *Journal of the Royal Statistical Society*, Ser. B, 40, 364 – 372.

—— (1983), "Bayesian "Confidence Intervals" for the Cross-Validated Smoothing Spline," *Journal of the Royal Statistical Society*, Ser. B, 45, 133–150.

—— (1986), "Partial and Interaction Splines for the Semiparametric Estimation of Functions of Several Variables," in *Computer Science and Statistics: Proceedings of the 18th Symposium on the interface*, ed. T.J. Boardman, American Statistical Association, pp. 75 – 80.

—— (1990), *Spline Models for Observational Data*, CBMS–NSF Regional Conference Series in Applied Mathematics, Vol. 59, SIAM.

Wahba, G. and Wendelberger, J. (1980), "Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross-Validation," *Monthly Weather Review*, 108, 1122–1145.