Partial and Interaction Spline Models for the Semiparametric Estimation of Functions of Several Variables.

Grace Wahba, University of Wisconsin–Madison

A partial spline model is a model for a response as a function of several variables, which is the sum of a "smooth" function of sevaral variables and a parametric function of the same plus possibly some other variables. Partial spline models in one and several variables, with direct and indirect data, with Gaussian errors and as an extension of GLIM to partially penalized GLIM models are described. Application to the modelling of change of regime in several variables is described. Interaction splines are introduced and described and their potential use for modelling nonlinear interactions between variables by semiparametric methods is noted. Reference is made to recent work in efficient computational methods.

1. Introduction

Partial spline models have proved to be interesting both from a practical and a theoretical point of view, partly because of their dual nature both as solutions to certain intuitively reasonable variational problems, and as Bayes estimates with certain parsimonious priors. In these proceedings we will attempt to give a quick rundown concerning some of their more interesting manifestations, and to report briefly on two new developments, first, the use of partial spline models to describe discontinuities or changes of regime, in two, three and higher dimensions, and, second, the idea of interaction splines for use in studying nonlinear interactions between variables semiparametrically.

2. Partial spline models - one splined variable

A response as a function of the variables $x, z_1, \ldots, z_k$ is modelled as

$$y_i = f(x(i)) + \sum_{j=1}^{p} \theta_j \Psi_j(x(i); z(i)) + \varepsilon_i \qquad (2.1a)$$

where

$$z(i) = (z_1(i), \ldots, z_k(i)) \qquad (2.1b)$$

the Ψ_j's are given parametric functions and the ε_i's are independent, zero mean Gaussian random variables with common (unknown) variance. The estimate $(f_\lambda, \theta_\lambda)$, where $\theta_\lambda = (\theta_{1\lambda}, \ldots, \theta_{p\lambda})$, is found as the minimizer, in an appropriate space, of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x(i)) - \sum_{j=1}^{p}\theta_j \Psi_j(x(i); z(i)))^2 +$$

$$\lambda J_m(f) \qquad (2.2a)$$

where

$$J_m(f) = \int_0^1 (f^{(m)}(x))^2 dx. \qquad (2.2b)$$

We have the following

Theorem: (Kimeldorf and Wahba (1971) - KW) Let $\Phi_1, \ldots, \Phi_m$ span the null space of J_m. If the design matrix for least squares regression on span $\Phi_1, \ldots, \Phi_m; \Psi_1, \ldots, \Psi_p$ is of full column rank, then there exists a unique minimizer $(f_\lambda, \theta_\lambda)$ for any $\lambda > 0$, and f_λ is a polynomial spline function.

The parameter λ as well as m can be choosen by generalized cross validation (GCV).

The appropriate function space here is the Sobolev space W_2^m, however, J_m (and W_2^m) can be replaced by any seminorm in a reproducing kernel (r. k.) Hilbert space of real valued functions on [0,1] provided that least squares regression onto the span of the null space of the seminorm is well defined - you get a Bayes estimate with the r. k. related to the prior covariance. Details may be found in KW and Wahba (1978) but we will not discuss the Bayesian aspect any further, other than to note that the prior behind J_m is the most parsimonious member of a large class of equivalent priors.

Partial spline models with one splined variable were introduced by several authors in different contexts, with some interesting applications, see Anderson and Senthilselvan (1982), Engle et al. (1983), Green, Jennison, and Seheult (1983), Shiller (1984).

3. Partial Spline Models - Several Splined Variables

Now, let the model be

$$y_i = f(x(i)) + \sum_{j=1}^{p} \theta_j \Psi_j(x(i); z(i)) + \varepsilon_i \qquad (3.1a)$$

where

$$x = (x_1, \ldots, x_d), \; x(i) = (x_1(i), \ldots, x_d(i)). \qquad (3.1b)$$

Again, we find f in an appropriate space to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x(i)) - \sum_{j=1}^{p}\theta_j \Psi_j(x(i); z(i)))^2 +$$

$$\lambda J_m(f) \qquad (3.2)$$

where now, we can use the "thin plate spline" penalty functional. For $d=2$, $m=2$, it is

$$J_m(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{x_1x_1}^2 + 2f_{x_1x_2}^2 + f_{x_2x_2}^2, \qquad (3.3)$$

and for arbitrary d it is

$$J_m(f) = \sum_{\alpha_1+\ldots+\alpha_d=m} \frac{m!}{\alpha_1! \cdots \alpha_d!} \times$$

$$\int \cdots \int \left[\frac{\partial^m f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}\right]^2 dx_1 \cdots dx_d. \qquad (3.4)$$

provided $2m > d$. The null space of J_m is the span of the $M = \binom{m+d-1}{d}$ monomials of total degree less than m, call them $\Phi_1, \ldots, \Phi_M$. Again, there will be a unique minimizer $(f_\lambda, \theta_\lambda)$ for every nonnegative λ if the design matrix for least squares regression on $\Phi_1, \ldots, \Phi_M; \Psi_1, \ldots, \Psi_p$ is of full column rank, and f_λ is a thin plate spline function.

Partial splines with several splined variables were introduced in Wahba (1984a), Wahba (1984b), Wahba (1985), and a discrete version has been proposed by Green, Jennison, and Seheult (1986). Transportable code (GCVPACK, Bates et al. (November 1985)) is available for fitting the partial spline models of (3.1)-(3.4) and computing the GCV estimate $\hat{\lambda}$ of λ. This code does well with up to around 400 data points on the VAX 11/750 in the Statistics Department at Madison. The work primarily depends on n, and not d, but, of course good estimates with large d will require large n. Diagnostics for splines (without the "partial" part) have been developed by Eubank (1986), it can be anticipated that this work will extend to partial spline models.

4. Indirect measurements

Let

$$g(x;z) = f(x) + \sum \theta_j \Psi_j(x;z), \qquad (4.1)$$

and now let

$$y_i = L_i g + \varepsilon_i \qquad (4.2)$$

where L_i is a bounded linear functional, for example:

$$L_i f = \int w_i(x;z) g(x;z) \pi dx \pi dz. \qquad (4.3)$$

This kind of data comes up in X-ray tomography, satellite tomography, stereology, and in other remote or indirect sensing problems in the physical and biological sciences. One finds f and θ to minimize:

$$\frac{1}{n} \sum (y_i - L_i f - \sum \theta_j L_i \Psi_j)^2 + \lambda J_m(f). \qquad (4.4)$$

The use of variants of (4.3), and (4.4) may also provide a good way to deal with heterogeneous aggregated economic data. For an application in stereology, see Nychka et al. (1984).

Data involving mildly nonlinear functionals can be accomodated - then

$$y_i = N_i g + \varepsilon_i \qquad (4.5a)$$

where

$$N_i g = \int \int w_i(x, z, g(x;z)) \pi dx \pi dz. \qquad (4.5b)$$

One finds f and θ to minimize

$$\frac{1}{n} \sum (y_i - N_i(f + \sum \theta_j \Psi_j))^2 + \lambda J_m(f). \qquad (4.6)$$

The minimization can be performed using basis functions and a Gauss-Newton iteration and λ chosen by GCV for nonlinear problems, see O'Sullivan and Wahba (1985).

5. Non Gaussian errors (semiparametric penalized GLIM models)

Here

$$g(x, z) = f(x) + \sum \theta_j \Psi_j(x;z) \qquad (5.1)$$

but

$$y_i \sim F_g.$$

For example :

$$y_i \sim Poisson \text{ with } \Lambda_i = e^{g(x(i);z(i))},$$

$$y_i \sim Binomial \text{ with } p_i/(1-p_i) = e^{g(x(i));z(i))},$$

etc. Here, one finds $f_\lambda, \theta_\lambda$ to minimize

$$L(f, \theta) + \lambda J_m(f) \qquad (5.2)$$

where L is the log likelihood. O'Sullivan (1983) and O'Sullivan, Yandell, and Raynor (1986) proposed numerical methods and a GCV for penalized GLIM models. See also Green and Yandell (1985), Silverman (1982), Cox and O'Sullivan (October, 1985), Leonard (1982). Further work on numerical methods for penalized GLIM and nonlinear indirect sensing problems is reported in this proceedings by Yandell.

6. Use of partial splines to model functions which are smooth except for specified discontinuities

Let $d=1$ and let

$$g(x;z) = f(x) + \theta |x - x^*|$$

that is, $\Psi_1(x;z) = |x - x^*|$. Then the partial spline estimate of g will have a jump in the first derivative at x^* of size 2θ. In two dimensions we may use a partial spline to model a jump in the first derivative with respect to x_2 along a given curve $x_{2*}(x_1)$: Let

$$\gamma(x) = \gamma(x_1, x_2) = |x_2 - x_2^*(x_1)|,$$

$$g(x;z) = f(x) + \theta(x_1)\gamma(x)$$

where θ may depend on x_1. Then

$$\left.\frac{\partial g}{\partial x_2}\right]_{x_2 = x_2(x_1)^*_-} - \left.\frac{\partial g}{\partial x_2}\right]_{x_2 = x_2(x_1)^*_+} = 2\theta(x_1).$$

If, for example

$$\theta(x_1) = \sum_{j=1}^{p} \theta_j q_j(x_1)$$

where the q_j's are given, then

$$\Psi_j(x;z) = q_j(x_1)\gamma(x).$$

This fits right into the partial spline setup, and GCVPACK may be used to compute the estimate. A generalization to $d=3$ with a jump in the first derivative with respect to x_3

along a surface $x_3^*(x_1,x_2)$ is straigtforward. For details, and a description of an application to the three dimensional modelling of the tropopause in the atmosphere and the thermocline in the ocean, see Shiau, Wahba, and Johnson (Dec. 1985).

7. Linear inequality constraints

Expressions (2.2), (3.2), (4.4), etc. can be minimized subject to finite families of linear inequality constraints. See Villalobos and Wahba (March 1985).

8. Main effects and interaction splines

The thin plate spline is defined on Euclidean d space for any d with $2m-d>0$, provided there are enough data points for mth degree polynomial regression, but unless there are very large data sets, in many applications will be desireable to reduce the amount of structure involved. Several authors have suggested modelling f as a linear combination of functions of one variable, that is,

$$f(x)=f_0+\sum_{al=1}^{d} f_\alpha(x_\alpha),$$

where $x=(x_1,\ldots,x_d)$, and $\int_0^1 f_\alpha(x_\alpha)dx_\alpha=0$. (Note the switch to the unit cube.) See Friedman, Grosse, and Stuetzle (1983), Stone (1985), Burman (June, 1985). We have been working on generalizations of this idea, whereby f is modelled successively as linear combinations of functions of one variable, functions of one and two variables, functions of one, two and three variables, etc. The resulting estimates may be called main effects splines, first order interaction splines, second order interaction splines, etc., by analogy with analysis of variance. We consider here two quite different but interesting penalty functionals which we will refer to as TEPR (for "tensor product"), and THPL (for "thin plate"). We will briefly sketch some early results of some work in progress, by describing the simplest examples.

The main ideas are most easily explained by first considering only spaces of periodic functions on the unit d-dimensional hypercube, that satisfy certain linear equality or boundary condtions, and then removing these conditions. Let $\phi_\nu(x_j)=\cos 2\pi\nu x_j$ or $\sin 2\pi\nu x_j$ (with some abuse of notation), and let $\theta_0=1, \theta_\nu=2\pi\nu, \nu>0$, and let H_{TEPR}^{per} and H_{THPL}^{per} be, respectively, the collections of all functions f of the form

$$f(x_1,\ldots,x_d)=$$

$$\sum_{\nu_1,\ldots,\nu_d=0}^{\infty} c_{\nu_1\cdots\nu_d}\phi_{\nu_1}(x_1)\cdots\phi_{\nu_d}(x_d) \qquad (8.1)$$

with

$$\sum_{\nu_1,\ldots,\nu_d=0}^{\infty}[\theta_{\nu_1}\cdots\theta_{\nu_d}]^{2m}c^2_{\nu_1\cdots\nu_d}<\infty, \quad H_{TEPR}^{per} \qquad (8.2)$$

or

$$\sum_{\nu_1,\ldots,\nu_d=0}^{\infty}[\theta_{\nu_1}^2+\cdots+\theta_{\nu_d}^2]^m c^2_{\nu_1\cdots\nu_d}<\infty \quad H_{THPL}^{per} (8.3)$$

It can be shown that H_{TEPR}^{per} will be a reproducing kernel hilbert space with (8.2) as squared norm for any $m>1/2$, and H_{THPL}^{per} will be a reproducing kernel space with the squared norm (8.3) for any $m>d/2$. These spaces are not equivalent, and reflect different ideas of what is "smooth". However, each can be written as the direct sum of 2^d orthogonal subspaces, namely, H_o, the $\binom{d}{1}$ "main effects" subspaces of the form

$$H_\alpha = span\ \{\phi_{\nu_\alpha}(x_\alpha), \nu_\alpha=1,2,\ldots\}\ \alpha=1,\ldots,d,$$

the $\binom{d}{2}$ first order interaction spaces of the form

$$H_{\alpha\beta}=span\ \{\phi_{\nu_\alpha}(x_\alpha)\phi_{\nu_\beta}(x_\beta), \nu_\alpha, \nu_\beta>0\},\ 1\le\alpha<\beta\le d,$$

and so on.

Letting

$$J_o(f)=[\int_0^1\cdots\int_0^1 f(x_1,\ldots,x_d)\prod_\alpha dx_\alpha]^2, \qquad (8.4)$$

the squared norm (8.2) on H_{TEPR}^{per} can be shown to be equal (in H_{TEPR}^{per}) to

$$J_o(f)+J^{THPL}(f), \qquad (8.5)$$

where

$$J^{THPL}(f)=\sum_{\alpha_1+\ldots+\alpha_d}\frac{m!}{\alpha_1!\cdots\alpha_d!}\times$$

$$\int_0^1\cdots\int_0^1\left[\frac{\partial^m f}{\partial x_1^{\alpha_1}\cdots\partial x_d^{\alpha_d}}\right]^2 dx_1\cdots dx_d \qquad (8.6)$$

is the thin plate penalty functional.

For lack of space we will not discuss the thin plate spaces further, but analyses similar to but slightly more complicated than those below can be carried out. In what follows, we will only consider the tensor product case and sub or superscripts TEPR are to be understood.

Let

$$J_\alpha(f)=\int_0^1 dx_\alpha[\int_0^1\cdots\int_0^1\frac{\partial^m f}{\partial x_\alpha^m}\prod_{\beta\ne\alpha}dx_\beta]^2 \qquad (8.7a)$$

$$J_{\alpha\beta}(f)=\int_0^1\int_0^1 dx_\alpha dx_\beta[\int_0^1\cdots\int_0^1\frac{\partial^{2m}f}{\partial x_\alpha^m\partial x_\beta^m}\prod_{\gamma\ne\alpha,\beta}dx_\gamma]^2 \quad (8.7b)$$

$$J_{1,\ldots,d}(f)=\int_0^1\cdots\int_0^1\left[\frac{\partial^{2md}f}{\partial x_1^m\cdots\partial x_d^m}\right]^2 dx_1\cdots dx_d. \quad (8.7c)$$

Then the squared norm (8.1) on H_{TEPR}^{per} can be shown to be equal to

$$J_o(f)+\sum_{\alpha=1}^{d} J_\alpha(f)+\sum_{\alpha<\beta} J_{\alpha\beta}(f)+\cdots+J_{1\ldots d}(f). \quad (8.8)$$

As an example, we will consider below $f \varepsilon H^{per}_{TEPR}$ which consists only of a mean, all d main effects and the first order interaction between x_1 and x_2. Thus f is of the form

$$f(x_1,\ldots,x_d)=f_0+\sum_{\alpha=1}^{d} f_\alpha(x_\alpha)+f_{12}(x_1,x_2), \quad (8.9)$$

where f_0 is a constant, $f_\alpha \varepsilon H_\alpha$, and $f_{12} \varepsilon H_{12}$. We can now define the periodic interaction smoothing spline as that function f_λ of the form (8.9) which minimizes

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-f(x(i)))^2+\lambda[\sum_{\alpha=1}^{d} J_\alpha(f_\alpha)+J_{12}(f_{12})], \quad \textbf{(8.10)}$$

where $x(i)=(x_1(i),\ldots,x_d(i))$.

Using Lemma 5.1 in KW it can be shown that there is a unique minimizer of (8.10) in $H_o \oplus \sum_\alpha H_\alpha \oplus H_{12}$. An explicit representation for it may be found using this lemma and the fact that the reproducing kernel $K(x,z)$ for $\sum_\alpha H_\alpha \oplus H_{12}$ is given by

$$K(x,z)=\sum_\alpha B_m(x_\alpha,z_\alpha)+B_m(x_1,z_1)B_m(x_2,z_2) \quad (8.11a)$$

where

$$B_m(s,t)=$$

$$\sum_{\nu=1}^{\infty}\theta_\nu^{-2m}[\cos 2\pi\nu s \cos 2\pi\nu t+\sin 2\pi\nu s \sin 2\pi\nu t] \quad (8.11b)$$

A closed form expression for B_m may be found in Craven and Wahba (1979). GCVPACK may be used to compute f_λ. In principle, $\sum_\alpha J_\alpha(f_\lambda)$ can be replaced by $\sum_\alpha w_\alpha J_\alpha(f_\lambda)$, where the w_α are positive weights, but problems concerning their estimation from the data have not been studied to date.

We will now sketch how to remove the rather restrictive periodicity conditions from H^{per}_{TEPR}. For g a function of one variable, let

$$L_0 g=\int_0^1 g(u)du \quad (8.12a)$$

$$L_\nu g=\int_0^1 g^{(\nu)}(u)du=g^{(\nu-1)}(1)-g^{(\nu-1)}(0), \quad (8.12b)$$

and let $L_{\nu(x_\alpha)}f$ mean L_ν applied to f as a function of x_α. Then $L_{\nu(x_\alpha)}f=0$ for $\nu=0,1,\ldots,m$, $\alpha=1,2,\ldots,d$, any f in H^{per}_{TEPR}. Now, it can be shown that H_α is that subspace of the Sobolev space

$$W_2^m[0,1]=\{g: g, g',\ldots,g^{(m-1)} abs. cont., g^{(m)} \varepsilon L_2\}$$

of co-dimension $m+1$ which satisfies the $m+1$ conditions $L_\nu g=0$, $\nu=0,1,\ldots,m$. Let $k_\nu=\frac{b_\nu}{\nu!}$, $\nu=0,1,\ldots,m$, where the b_ν are the Bernoulli polynimials, we have $L_\nu k_\mu=0, \mu\neq\nu, L_\nu k_\nu=1, \mu,\nu=0,1,\ldots,m$, and thus k_ν is not in H_α. Let $W^0=span\{k_0,\ldots,k_{m-1}\}$ and let W^1 be isomorphic to $H_\alpha \oplus \{k_m\}$. Then it can be shown that W_2^m endowed with the inner product

$$<g,h>_{W_2^m}=\sum_\nu L_\nu g L_\nu h+\int_0^1 g^{(m)}(u)h^{(m)}(u)du \quad (8.13)$$

satisfies

$$W_2^m=W^0 \oplus W^1. \quad (8.14)$$

Letting $g \varepsilon W_2^m$ with $g=g_0+g_1, g_0 \varepsilon W_0, g_1 \varepsilon W_1$ we can call g the polynomial part of g, and g_1 the "smooth" part. Now let

$$H_{THPL}=W_2^m\otimes\cdots\otimes W_2^m \quad d \text{ times} \quad (8.15)$$

$$=(W^0\oplus W^1)\otimes\cdots\otimes(W^0\oplus W^1).$$

$$=(\prod_{\alpha=1}^{d} W_\alpha^0)\oplus(\sum_{\alpha=1}^{d} W_\alpha^1\otimes\prod_{\substack{\beta=1\\ \beta\neq\alpha}} W_\beta^0)\oplus$$

$$(\sum_{\alpha<\beta} W_\alpha^1\otimes W_\beta^1\otimes\prod_{\gamma\neq\alpha,\beta} W_\alpha^0)\oplus\cdots\oplus$$

$$(\prod_{\alpha=1}^{d} W_\alpha^1),$$

where the Greek subscripts make explicit which variables are involved. We can now identify the "polynomial" subspace

$$H_0=\prod_{\alpha=1}^{d} W_\alpha^0,$$

the main effects subspaces

$$H_\alpha=W_\alpha^1\otimes\prod_{\substack{\beta=1\\ \beta\neq\alpha}}^{d} W_\beta^0, \ \alpha=1,\ldots,d,$$

the first order interaction spaces

$$H_{\alpha\beta}=W_\alpha^1\otimes W_\beta^1\otimes\prod_{\gamma\neq\alpha,\beta} W_\alpha^0,$$

etc.

The induced tensor product inner product in H_{TEPR} is a natural extension of the inner product of (8.7) and (8.8). Letting J_α be the induced norm on H_α, etc., we can now seek f_λ in the new, non periodic version of, for example $H_0\oplus\sum_\alpha H_\alpha\oplus H_{12}$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-f_\lambda x(i))^2+\lambda[\sum_\alpha J_\alpha(f_\alpha)+J_{12}(f_{12})]. \quad (8.16)$$

Existence and uniqueness for any $\lambda>0$ can be shown via Lemma 5.1 in KW provided the design points $x(i), i=1,\ldots,n$ are such that least squares regression in

H_0 is unique. The reproducing kernels for the various subspaces then follow:- The r. k.'s R_0 and R_1 for W^0 and W^1 can be shown to be

$$R_0(u,v) = \sum_{v=0}^{m-1} k_v(u) k_v(v),$$

$$R_1(u,v) = k_m(u)k_m(v) + B_m(u,v),$$

and the r. k. for H_{TEPR} with the inner product induced by (8.13) is

$$\prod_{\alpha=1}^{d} (R_0(x_\alpha, z_\alpha) \oplus R_1(x_\alpha, z_\alpha)),$$

thus, for example the r. k. for $\sum_\alpha H_\alpha \oplus H_{12}$ is now

$$Q(x,z) = \sum_\alpha R_1(x_\alpha, z_\alpha) \prod_{\beta \neq \alpha} R_0(x_\beta, z_\beta) +$$

$$R_1(x_1, z_1) R_1(x_2, z_2) \prod_{\beta \neq 1,2} R_0(x_\beta, z_\beta).$$

Given the r. k. an explicit representation for f_λ can be given, and, again GCVPACK can be used to calculate f_λ. For $m=1$, $R_0(x_\alpha, z_\alpha) = 1$, H_0 is one dimensional as before, and we only replace B_m in the discussion of periodic spaces by R_1 and the same expressions hold. For $m>1$, a typical element of H_α with, say $\alpha = 1$ is now of the form

$$f(x_1, \ldots, x_d) =$$

$$\sum_{v_1, \ldots, v_d = 0}^{m-1} f_{v_2 \cdots v_d}(x_1) k_{v_2}(x_2) \cdots k_{v_d}(x_d). \quad (8.18)$$

The $v_2 = \cdots = v_d = 0$ term depends only on x_1 but the other terms do depend on the other variables albeit in a parametric (i. e. polynomial) way. The case $m=2$ is probably of special interest, then x_β with $\beta \neq \alpha$ enters at most linearly in functions in H_α.

There are now many interesting questions. Some of the major ones are - the development of good methods for choosing which interactions to include (GCV?), numerical methods for vary large data sets, methods for interpreting the results, development of confidence intervals, and so on.

Acknowledgements

This research supported in part by NASA under Contract NAG5-316 and by NSF under Grant DMS 8404970.

References

Anderson, J. A., and Senthilselvan, A. (1982), "A Two-Step Regression Model for Hazard Functions," *Appl. Statist.*, 31, 44-51.

Bates, D. M., Lindstrom, M. J., Wahba, G., and Yandell, B. (November 1985) "GCVPACK - Routines for Generalized Cross Validation." Technical Report 775, Dept. of Statistics, University of Wisconsin-Madison.

Burman, P. (June, 1985), *Estimation of Generalized Additive Models*, Rutgers University. (manuscript)

Cox, D. D., and O'Sullivan, F. (October, 1985), *Analysis of Penalized Likelihood-Type Estimayors with Application to Generalized Smoothing in Sobolev Spaces*, manuscript.

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.

Engle, R., Granger, C., Rice, J., and Weiss, A. (1983) "Nonparametric Estimates of the Relation Between Weather and Electricity Demand.." Discussion paper 83-17, San Diego: Dept. of Economics, University of California,.

Eubank, R.L. (1986), "Diagnostics for Smoothing Splines," *J. Roy. Stat. Soc. Ser. B*, 47. (to appear)

Friedman, J. H., Grosse, E., and Stuetzle, W. (1983), "Multidimensional Additive Spline Approxmation," *SIAM J. Sci.Stat. Comput.*, 4, 291-301.

Green, P., Jennison, C., and Seheult, A. (1983), "Comments to Nearest Neighbour (NN) Analysis of Field Experiments by Wilkinson, Et. al.," *Journal of the Royal Statistical Society, Ser. B*, 45, 193-195.

Green, P. J., and Yandell, B. S. (1985) "Semi-Parametric Generalized Linear Models." Technical Report#2847, Math. Research Center, U. of Wisconsin.

Green, P., Jennison, C., and Seheult, A. (1986), "Analysis of Field Experiments by Least Squares Smoothing," *Journal of the Royal Statistical Society, Ser. B*, 47.

Kimeldorf, G., and Wahba, G. (1971), "Some Results on Tchebycheffian Spline Functions," *J. Math. Anal. Applic.*, 33, 82-95.

Leonard, T. (1982) "An Empirical Bayesian Approach to the Smooth Estimation of Unknown Functions.." Technical Report 2339, Mathematics Research Center, Univ. of Wisconsin.

Nychka, D., Wahba, G., Goldfarb, S., and Pugh, T. (1984), "Cross-Validated Spline Methods for the Estimation of Three Dimensional Tumor Size Distributions from Observations on Two Dimensional Cross Sections," *J. Am. Stat. Assoc.*, 79, 832-846.

O'Sullivan, F. (1983), *The Analysis of Some Penalized Likelihood Estimation Schemes*, University of Wisconsin-Madison, Statistics Dept.. (Technical Report #726)

O'Sullivan, F., and Wahba, G. (1985), "A Cross Validated Bayesian Retrieval Algorithm for Non-Linear Remote Sensing Experiments," *J. Comput. Physcs*, 59, 441-455.

O'Sullivan, F., Yandell, B., and Raynor, W. (1986), "Automatic Smoothing of Regression Functions in Generalized Linear Models," *J. Am. Stat. Soc.*, 81, 96-103.

Shiau, J., Wahba, G., and Johnson, D. R. (Dec. 1985) "Partial Spline Models for the Inclusion of Tropopause and Frontal Boundary Information in Otherwise Smooth Two and Three Dimensional Objective Analysis." Technical Report # 777 University of Wisconsin-Madison Statistics Dept.

Shiller, R. J. (1984), "Smoothness Priors and Nonlinear Regression," *J. Am. Stat. Soc.*, 79, 609-615.

Silverman, B. (1982), "On the Estimation of a Probability Density Function by the Maximum Penalized Likelihood Method," *Ann. Statist.*, 10, 795-810.

Stone, C. J. (1985), "Additive Regression and Other Nonparametric Models," *Ann. Statist.*, 13, 689-705.

Villalobos, M., and Wahba, G. (March 1985), *Inequality Constrained Multivariate Smoothing Splines with Application to the Estimation of Posterior Probabilities*, Dept. of Statistics, University of Wisconsin-Madison. (Technical Report No. 756)

Wahba, G. (1978), "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression," *J. R. Statist. Soc. B*, 40, 364-372.

Wahba, G. (1984a), "Cross Validated Spline Methods for the Estimation of Multivariate Functions from Data on Functionals," in *Statistics: An Appraisal, Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory*, eds. H. A. David, and H. T. David Iowa State University Press.

Wahba, G. (1984b), "Partial Spline Models for the Semiparametric Estimation of Functions of Several Variables," in *Statistical Analyses of Time Series*, Tokyo: Institute of Statistical Mathematics, 319-329. (Proceedings of the Japan U. S. Joint Seminar)

Wahba, G. (1985), "Comments to Peter J. Huber, Projection Pursuit," *Ann. Statist.*, 13, 518-521.