The Annals of Statistics
1986, Vol. 14, No. 3, 934–953

NONPARAMETRIC BAYESIAN REGRESSION

By Daniel Barry

University College, Cork

It is desired to estimate a real valued function F on the unit square having observed F with error at N points in the square. F is assumed to be drawn from a particular Gaussian process and measured with independent Gaussian errors. The proposed estimate is the Bayes estimate of F given the data. The roughness penalty corresponding to the prior is derived and it is shown how the Bayesian technique can be regarded as a generalisation of variance components analysis. The proposed estimate is shown to be consistent in the sense that the expected squared error averaged over the data points converges to zero as $N \to \infty$. Upper bounds on the order of magnitude of the expected average squared error are calculated. The proposed technique is compared with existing spline techniques in a simulation study. Generalisations to higher dimensions are discussed.

1. Introduction and summary. Let $(y_i, \mathbf{x}_i)$, $i = 1, 2, \dots, N$, satisfy

$$y_i = F(\mathbf{x}_i) + e_i,$$

where $\mathbf{x}_i \in [0,1] \times [0,1]$ for each i, F is a fixed but unknown regression function, and the errors $\{e_i\}$ are uncorrelated with mean zero and variance v. This paper concerns estimation of F. The classic example of this situation is where $\mathbf{X} = (X_1, X_2)$ specifies coordinates on a plane and Y is a measure such as height above sea level. Our aim is to use the data to construct a map.

Wahba (1978) considers the above problem when $x_i \in [0,1]$, $i = 1, 2, \dots, N$. Later in Wahba (1979) she extended the technique to cover estimation in higher dimensions. Here we consider another possible generalisation to two dimensions of Wahba's work and claim that it fits more satisfactorily into the Bayesian formulation described in Wahba (1978).

Motivated by a decomposition often used in two-way analysis of variance we can write

$$F(x_1, x_2) = \mu + \alpha(x_1) + \beta(x_2) + \gamma(x_1, x_2),$$

where

$$\mu = \int_0^1 \int_0^1 F(u, v)\, du\, dv,$$

$$\alpha(x_1) = \int_0^1 F(x_1, v)\, dv - \mu,$$

$$\beta(x_2) = \int_0^1 F(u, x_2)\, du - \mu,$$

$$\gamma(x_1, x_2) = F(x_1, x_2) - \alpha(x_1) - \beta(x_2) - \mu.$$

Received January 1984; revised November 1985.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62J05, 62M99.
Key words and phrases. Bayes estimate, Brownian sheet, roughness penalty, consistency.

A prior for F is constructed by putting independent priors on μ, α, β, and γ as follows:

(i) $\mu \sim N(0, v_0)$.

(ii) $\alpha(x_1) \sim Z_1(x_1) - \int_0^1 Z_1(u)\,du$ where Z_1 is a Brownian motion with variance v_1.

(iii) $\beta(x_2) \sim Z_2(x_2) - \int_0^1 Z_2(u)\,du$ where Z_2 is a Brownian motion with variance v_2.

(iv) $\gamma(x_1,x_2) \sim Z(x_1,x_2) - \int_0^1 Z(x_1,u)\,du - \int_0^1 Z(u,x_2)\,du + \int_0^1\int_0^1 Z(u,v)\,du\,dv$, where Z is a Brownian sheet with variance v_{12}.

To complete the specification of the probability model required for a Bayesian analysis we assume (v) given F, the data $\{y_i\}$ are independent with

$$y_i \sim N[F(x_i), v], \qquad i = 1, 2, \dots, N.$$

The proposed estimate of F is then the limit of the Bayes estimate of F as $v_0 \to \infty$, i.e.,

$$\hat{F}(x_1, x_2) = \lim_{v_0 \to \infty} E_{v_0}\{F(x_1, x_2) | y_1, y_2, \dots, y_N\},$$

where E_{v_0} is expectation with respect to the posterior density resulting from the probability model defined in (i) to (v) above.

In Section 2 we indicate how the above prior came about as a generalisation of the one-dimensional priors described in Wahba (1978). Section 3 demonstrates that the suggested Bayesian analysis corresponds to choosing $\hat{F}$ to minimise

$$\sum_{i=1}^{N} (y_i - F(\mathbf{x}_i))^2 + P(F),$$

where

$$P(F) = c_1 \int_0^1 \left[\int_0^1 F_1(x_1, v)\,dv \right]^2 dx_1 + c_2 \int_0^1 \left[\int_0^1 F_2(u, x_2)\,du \right]^2 dx_2 + c_{12} \int_0^1 \int_0^1 F_{12}(u, v)^2\,du\,dv,$$

F_1, F_2 are partial derivatives, F_{12} is the second-order mixed partial derivative, and $c_1 = v/v_1$, $c_2 = v/v_2$, $c_{12} = v/v_{12}$. The possibility of regarding the technique as a generalisation of variance components analysis is examined in Section 4.

Asymptotic properties are considered in Sections 5 and 6. These properties depend only on the assumption that the errors are uncorrelated with mean zero and constant variance. The symbol E will be used throughout to denote expectation with respect to these assumptions. Thus the expectations have a frequentist interpretation since they can be interpreted without reference to the prior; from a Bayesian viewpoint they are expectations conditional on F. In Section 5 we prove:

THEOREM. *Define*

$$R(\hat{F}, F) = \frac{1}{N} \sum_{i=1}^{N} (\hat{F}(\mathbf{x}_i) - F(\mathbf{x}_i))^2.$$

Then, under certain smoothness assumptions on F,

$$ER(\hat{F}, F) \le \frac{1}{N}P(F) + \left(\frac{v}{4}\right)\left[\frac{12}{N} + \frac{v_1}{v} + \frac{v_2}{v} + \frac{v_{12}}{v}\right].$$

Hence $ER(\hat{F}, F) \to 0$ as $N \to \infty$ provided

$$v_1 \to 0, \quad Nv_1 \to \infty, \quad v_2 \to 0, \quad Nv_2 \to \infty, \quad v_{12} \to 0, \quad Nv_{12} \to \infty.$$

In Section 6 we consider data on a grid and prove:

THEOREM. *Suppose we have $N = mn$ observations on a grid*

$$\{(x_{1i}, x_{2j}): 1 \le i \le m, 1 \le j \le n\}.$$

Then, under certain smoothness assumptions on F,

$$ER(\hat{F}, F) \le \frac{1}{N}P(F) + O\left[\left(\frac{B_1(v_1 + v_{12})}{n}\right)^{1/2}\right]$$
$$+ O\left[\left(\frac{B_2(v_2 + v_{12})}{m}\right)^{1/2}\right] + O\left[(B_1 B_2 v_{12})^{1/2} \log(B_1 B_2 v_{12})\right],$$

where $B_1 = \max(x_{1i} - x_{1i-1})$, $B_2 = \max(x_{2j} - x_{2j-1})$.

This theorem provides more precise rates of convergence in the grid case. The choice of values for c_1, c_2, and c_{12} in a particular application is difficult and is typical of all nonparametric estimation techniques. In the simulation study described in this paper Wahba's method of generalised cross validation [see Craven and Wahba (1979)] was used to choose values for c_1, c_2, and c_{12}. It remains to be shown that the values so obtained have the orders of magnitude suggested by the asymptotic calculations.

Wahba's two-dimensional regression technique is described in Section 7 and is compared with the Bayesian technique in a simulation study reported in Section 8. Finally, generalisations of the Bayesian technique to higher dimensions are described in Section 9.

2. The Bayes estimate. Wahba (1978) considers the problem of estimating a regression function F: $[0,1] \to R$ given data (x_i, y_i), $i = 1, 2, \dots, N$, with

$$y_i = F(x_i) + e_i,$$

where $\{e_i\}$ are iid $N(0, v)$. A prior for F is constructed by writing

$$F(x) = \mu + \alpha(x),$$

where $\mu = F(0)$, $\alpha(x) = F(x) - F(0)$ and then putting independent priors on μ and α as follows:

(i) $\mu \sim N(0, v_0)$.
(ii) $\alpha(x) \sim Z(x)$ where Z is a Brownian motion with variance v_1, i.e., Z is a

mean-zero Gaussian process with

$$\operatorname{Cov}[Z(x), Z(y)] = v_1 \min(x, y).$$

The estimator $\hat{F}$ is defined by

$$\hat{F}(x) = \lim_{v_0 \to \infty} E_{v_0}\{F(x) | y_1, y_2, \dots, y_n\},$$

where E_{v_0} is expectation with respect to the posterior density generated by the prior and the normality assumptions on the errors $\{e_i\}$. Wahba shows that this procedure is equivalent to choosing F absolutely continuous to minimise

$$\sum (y_i - \hat{F}(x_i))^2 + c\int_0^1 \hat{F}'(x)^2 \, dx,$$

where $c = v/v_1$. She further describes the equivalence between roughness penalties of the form $\int_0^1 \hat{F}^{(m)}(x)^2 \, dx$ and corresponding Bayesian priors which for $m > 1$ are integrated Wiener processes [see Shepp (1966)]. Here we generalise the $m = 1$ prior to the situation where we have data in two dimensions and follow Wahba in defining the estimator.

One possible generalisation of Brownian motion to two dimensions is called a Brownian sheet. This is a Gaussian process Z indexed by $\{(x_1, x_2): x_1 \geq 0, x_2 \geq 0\}$ and satisfying

$$EZ(x_1, x_2) = 0,$$

$$\operatorname{Cov}[Z(x_1, x_2), Z(y_1, y_2)] = \min(x_1, y_1)\min(x_2, y_2).$$

See Zimmerman (1972).

The one-dimensional prior was based on the decomposition

$$F(x) = F(0) + (F(x) - F(0))$$

with a Brownian motion prior on the term $F(x) - F(0)$. Consider a similar decomposition here,

$$F(x_1, x_2) = F(0,0) + (F(x_1, x_2) - F(0,0)),$$

with a Brownian sheet prior on $F(x_1, x_2) - F(0,0)$. This is not satisfactory since with probability one a Brownian sheet is constant along the lines $x_1 = 0$ and $x_2 = 0$. We do not want to impose such conditions on our estimator.

Alternatively we might consider the following four-way decomposition:

$$F(x_1, x_2) = F(0,0) + [F(x_1, 0) - F(0,0)] + [F(0, x_2) - F(0,0)] + [F(x_1, x_2) + F(0,0) - F(x_1, 0) - F(0, x_2)].$$

We could then put independent priors on each part: $F(0,0) \sim N(0, v_0)$, Brownian motions on $[F(x_1, 0) - F(0,0)]$ and $[F(0, x_2) - F(0,0)]$ and a Brownian sheet on $(F(x_1, x_2) + F(0,0) - F(x_1, 0) - F(0, x_2))$. The problem with this approach however is that it imposes different smoothness conditions on the function values along the axes through zero.

This can be seen by applying the argument to be given in Section 3 to this Bayesian model to show that the Bayes estimate in this case corresponds to the use of a roughness penalty of the form

$$\beta_1\int_0^1 F_1(x,0)^2\,dx + \beta_2\int_0^1 F_2(0,y)^2\,dy + \beta_{12}\int_0^1\int_0^1 F_{12}(x,y)^2\,dy\,dx,$$

where F_1, F_2 are partial derivatives, F_{12} is the second-order mixed partial derivative, and $\beta_1, \beta_2, \beta_{12}$ are constants depending on the error variance v and the variances associated with the elements of the prior specification. There is rarely a prior justification for this.

To avoid this difficulty we propose the following decomposition:

$$F(x_1,x_2) = \mu + \alpha(x_1) + \beta(x_2) + \gamma(x_1,x_2),$$

where

$$\mu = \int\int F(u,v)\,du\,dv \quad \text{all integrals have range } [0,1],$$

$$\alpha(x_1) = \int F(x_1,v)\,dv - \int\int F(u,v)\,du\,dv,$$

$$\beta(x_2) = \int F(u,x_2)\,du - \int\int F(u,v)\,du\,dv,$$

$$\gamma(x_1,x_2) = F(x_1,x_2) - \int F(x_1,v)\,dv - \int F(u,x_2)\,du + \int\int F(u,v)\,du\,dv.$$

We proceed to put independent priors on μ, α, β, and γ as follows:

1. $\mu \sim N(0, v_0)$.
2. $\alpha(x_1) \sim Z_1(x_1) - \int Z_1(u)\,du$, where $Z_1(u)$ is Brownian motion with variance v_1.
3. $\beta(x_2) \sim Z_2(x_2) - \int Z_2(v)\,dv$, where $Z_2(v)$ is Brownian motion with variance v_2.
4. $\gamma(x_1,x_2) \sim Z(x_1,x_2) - \int Z(u,x_1)\,du - \int Z(x_2,v)\,dv + \int\int Z(u,v)\,du\,dv$, where $Z(u,v)$ is a Brownian sheet with variance v_{12}.

The following results may be easily shown:

1. $\mathrm{Cov}(\alpha(x),\alpha(y)) = v_1[\min(x,y) - x - y + (x^2+y^2)/2 + \frac{1}{3}]$
   $= v_1 h(x,y)$, say.
2. $\mathrm{Cov}(\beta(x),\beta(y)) = v_2 h(x,y)$.
3. $\mathrm{Cov}(\gamma(x_1,x_2),\gamma(y_1,y_2)) = v_{12}h(x_1,y_1)h(x_2,y_2)$.

In summary, therefore, the prior specification is the same as the distribution of the process

$$\mu + R(x_1,x_2),$$

where $\mu \sim N(0,v_0)$, and $R(u,v)$ is a mean zero Gaussian process with

$$\begin{aligned}\mathrm{Cov}(R(x_1,x_2),R(y_1,y_2)) &= v_1 h(x_1,y_1) + v_2 h(x_2,y_2) \\ &\quad + v_{12}h(x_1,y_1)h(x_2,y_2) \\ &= Q(\mathbf{x},\mathbf{y}), \quad \text{say.}\end{aligned}$$

We now assume that we have observations $(\mathbf{x}_i, y_i)$, $i = 1,2,\dots,N$, with

$$\mathbf{x}_i = (x_{1i}, x_{2i}) \in [0,1]\times[0,1]$$

and

$$y_i = F(\mathbf{x}_i) + e_i$$

with $\{e_i\}$ iid $N(0, v)$. The proposed estimator is defined by

$$F(x_1, x_2) = \lim_{v_0 \to \infty} E_{v_0}\{F(x_1, x_2)|y_1, y_2, \dots, y_N\},$$

where E_{v_0} is expectation with respect to the posterior density generated by the above prior and the normality assumptions on the error terms.

3. The roughness penalty. For a function F: $[0,1] \times [0,1] \to R$, let F_1 denote the partial derivative with respect to x_1, F_2 the partial derivative with respect to x_2, and F_{12} the second-order mixed partial derivative.

Let H consist of all functions F: $[0,1] \times [0,1] \to R$ satisfying the following conditions:

(i) F is absolutely continuous.
(ii) F_1 is an absolutely continuous function of x_2 for each x_1 in $[0,1]$.
(iii) F_2 is an absolutely continuous function of x_1 for each x_2 in $[0,1]$.
(iv) $F_{12} \in L^2[0,1]^2$, i.e., $\int_0^1\int_0^1 F_{12}^2(u,v)\,du\,dv < \infty$.

Then we have the following theorem.

THEOREM 3.1. *The Bayes estimate generated by the prior of Section* 2 *is the unique element in H minimising*

$$\sum_{i=1}^{N} (y_i - F(\mathbf{x}_i))^2 + P(F), \tag{3.1}$$

where

$$P(F) = c_1\int_0^1\left[\int_0^1 F_1(x_1, v)\,dv\right]^2 dx_1 + c_2\int_0^1\left[\int_0^1 F_2(u, x_2)\,du\right]^2 dx_2 + c_{12}\int_0^1\int_0^1 F_{12}(u,v)^2\,du\,dv$$

with $c_1 = v/v_1$, $c_2 = v/v_2$, *and* $c_{12} = v/v_{12}$.

PROOF. For $F, G \in H$ define

$$\begin{aligned}\langle F, G\rangle = &\left[\int\int F(u,v)\,du\,dv\right]\left[\int\int G(u,v)\,du\,dv\right]\\ &+a_1\int\left[\int F_1(x_1,u)\,du\right]\left[\int G_1(x_1,v)\,dv\right]dx_1\\ &+a_2\int\left[\int F_2(u,x_2)\,du\right]\left[\int G_2(v,x_2)\,dv\right]dx_2\\ &+a_{12}\int\int F_{12}(u,v)G_{12}(u,v)\,du\,dv,\end{aligned}$$

where $a_1 = 1/v_1$, $a_2 = 1/v_2$, and $a_{12} = 1/v_{12}$.

It can be shown that $\langle \cdot, \cdot \rangle$ is an inner product for H. The only difficulty is in proving that $\langle F, F \rangle = 0 \Rightarrow F = 0$. Now

$$\begin{aligned} \langle F, F \rangle = 0 &\Rightarrow \int\int F_{12}^2 = 0 \\ &\Rightarrow F_{12} = 0 \quad \text{a.e.} \\ &\Rightarrow F_1 \text{ is a function of } x_1 \text{ alone and } F_2 \text{ is a function of } x_2 \text{ alone.} \end{aligned}$$

Also

$$\langle F, F \rangle = 0 \Rightarrow \int \left[\int F_1(x_1, v)\, dv \right]^2 dx_1 = 0, \qquad F_1 = 0 \quad \text{a.e.}$$

Similarly $\langle F, F \rangle = 0 \Rightarrow F_2 = 0$ a.e. and hence $\langle F, F \rangle = 0 \Rightarrow F$ is constant. But

$$\begin{aligned} \langle F, F \rangle = 0 &\Rightarrow \int\int F = 0 \\ &\Rightarrow F = 0 \quad \text{a.e.} \end{aligned}$$

Since F is absolutely continuous we have $F = 0$. For $\mathbf{x}, \mathbf{y} \in [0,1] \times [0,1]$ define

$$\phi_0(\mathbf{x}) = 1,$$

$$Q(\mathbf{x}, \mathbf{y}) = v_1 h(x_1, y_1) + v_2 h(x_2, y_2) + v_{12} h(x_1, y_1) h(x_2, y_2),$$

where

$$h(x, y) = \min(x, y) - (x + y) + \tfrac{1}{2}(x^2 + y^2) + \tfrac{1}{3}.$$

Then

(1) $$\phi_0 \in H; \qquad Q(\mathbf{x}, \cdot) \in H \quad \text{for all } \mathbf{x},$$

(2) $$\langle \phi_0, Q(\mathbf{x}, \cdot) \rangle = 0 \quad \text{for all } \mathbf{x},$$

(3) $$\langle F, \phi_0 + Q(\mathbf{x}, \cdot) \rangle = F(\mathbf{x}) \quad \text{for all } F \in H.$$

(3) follows easily upon noting that

$$\int_0^1 h(u, v)\, dv = 0 \quad \text{for all } u$$

and that for any absolutely continuous function

$$\int_0^1 \psi'(y) \frac{\partial}{\partial y} h(x, y)\, dy = \psi(x) - \int_0^1 \psi(y)\, dy.$$

It follows from (1)–(3) that

$$\langle Q(\mathbf{y}, \cdot), Q(\mathbf{x}, \cdot) \rangle = Q(\mathbf{y}, \mathbf{x}).$$

For any function $F \in H$ we can write

$$F(\mathbf{x}) = a\phi_0(\mathbf{x}) + \sum_{i=1}^{N} b_i Q(\mathbf{x}_i, \mathbf{x}) + e(\mathbf{x}),$$

where

$$\langle \phi_0, e \rangle = 0,$$

$$\langle Q(\mathbf{x}_i, \cdot), e \rangle = 0, \qquad i = 1, 2, \ldots, N.$$

Hence

$$F(\mathbf{x}_i) = \langle F, \phi_0 + Q(\mathbf{x}_i, \cdot)\rangle$$
$$= a + \sum_{j=1}^{N} b_j Q(\mathbf{x}_j, \mathbf{x}_i).$$

Also

$$P(F) = v\langle F, F\rangle - v\left[\int_0^1\int_0^1 F(u, v)\, du\, dv\right]^2$$
$$= v\sum\sum b_i b_j Q(\mathbf{x}_i, \mathbf{x}_j) + v\langle e, e\rangle.$$

Thus to minimise (3.1) we choose $e = 0$ and $a, \mathbf{b} = (b_1, b_2, \dots, b_N)$ to minimise

$$(\mathbf{y} - a1 - \mathbf{b}Q)'(\mathbf{y} - a1 - \mathbf{b}Q) + v\mathbf{b}Q\mathbf{b}',$$

where 1 is a vector of ones

$$Q = \big(Q(\mathbf{x}_i, \mathbf{x}_j)\big), \qquad i, j = 1, 2, \dots, N.$$

As in Lemma 5.1 of Kimeldorf and Wahba (1971) the minimising values are given by

$$a = (1'M^{-1}1)^{-1}1'M^{-1}\mathbf{y},$$
$$\mathbf{b} = M^{-1}\left[I - 1(1'M^{-1}1)^{-1}1'M^{-1}\right]\mathbf{y},$$

where $M = Q + vI$. That this formula also gives the Bayes estimate follows as in Theorem 1 of Wahba (1978). □

NOTES. 1. The form of the roughness penalty indicates how scale changes in the X-variables can be taken into account by adjusting the smoothing parameters, e.g., if we rescale X_1 to $W_1 = \alpha + \beta x_1$ then adjusting c_1 to c_1/β and c_{12} to c_{12}/β leaves the roughness penalty unchanged.

2. Let $\hat{\mathbf{F}} = (\hat{F}(\mathbf{x}_1), \hat{F}(\mathbf{x}_2), \dots \hat{F}(\mathbf{x}_N))$. Then ∃ a matrix A with

$$\hat{F} = A\mathbf{y}$$

and A is given by

$$A = 1(1'M^{-1}1)^{-1}1'M^{-1} + QM^{-1}\left[I - 1(1'M^{-1}1)^{-1}1'M^{-1}\right]$$
$$= QM^{-1} + vM^{-1}1(1'M^{-1})^{-1}1'M^{-1},$$

where

$$1 = (1, 1, \dots, 1)', \qquad N \times 1,$$
$$M = Q + vI, \qquad N \times N,$$
$$Q = \big(Q(\mathbf{x}_i, \mathbf{x}_j)\big), \qquad N \times N.$$

Evaluation of $\mathbf{F}$ therefore involves inverting the $N \times N$ matrix M.

3. The completeness of H is left open; the method of proof used in Wahba (1978) does not depend on it.

4. Data on a grid in two dimensions. In this section we derive a representation for the estimator $\hat{F}$ when observations are taken on a grid in two dimensions.

THEOREM 4.1. *Suppose we have* $N = mn$ *observations on a grid*

$$\{(x_{1i}, x_{2j}): 1 \le i \le m, 1 \le j \le n\},$$

where

$$0 \le x_{11} < x_{12} < \cdots < x_{1m} \le 1,$$
$$0 \le x_{21} < x_{22} < \cdots < x_{2n} \le 1.$$

Let $F_{ij} = F(x_{1i}, x_{2j})$ *and define*

$$\mathbf{F} = (F_{11}, F_{21}, \dots, F_{m1}, F_{12}, F_{22}, \dots, F_{m2}, \dots, F_{1n}, F_{2n}, \dots, F_{mn})'$$

and define $\mathbf{y}$ *similarly. Let*

$$\mathbf{a}_1 = \left(\tfrac{1}{2}(x_{11} + x_{12}), \tfrac{1}{2}(x_{13} - x_{11}), \dots, 1 - \tfrac{1}{2}(x_{1m} + x_{1m-1})\right)',$$
$$\mathbf{b}_1 = \left(\tfrac{1}{2}(x_{21} + x_{22}), \tfrac{1}{2}(x_{23} - x_{21}), \dots, 1 - \tfrac{1}{2}(x_{2n} + x_{2n-1})\right)'.$$

For $1 \le i \le m$ *define*

$$\alpha(\mathbf{F}, i) = \sum_{j=1}^{n} a_{1j} F_{ij}.$$

For $1 \le j \le n$ *define*

$$\beta(F, j) = \sum_{i=1}^{m} b_{1i} F_{ij}.$$

Then the Bayes estimate $\hat{\mathbf{F}}$ *is chosen to minimise*

$$(\mathbf{y} - \mathbf{F})'(\mathbf{y} - \mathbf{F}) + d_1 \sum_{i=2}^{m} \frac{[\alpha(F, i) - \alpha(F, i-1)]^2}{x_{1i} - x_{1i-1}}$$
$$+ d_2 \sum_{j=2}^{n} \frac{[\beta(F, j) - \beta(F, j-1)]^2}{x_{2j} - x_{2j-1}}$$
$$+ d_{12} \sum_{i=2}^{m} \sum_{j=2}^{n} \frac{(F_{ij} + F_{i-1j-1} - F_{ij-1} - F_{i-1j})^2}{(x_{1i} - x_{1i-1})(x_{2j} - x_{2j-1})},$$

where

$$d_1 = v/(v_1 + v_{12}\mathbf{b}_1' Q_2 \mathbf{b}_1),$$
$$d_2 = v/(v_2 + v_{12}\mathbf{a}_1' Q_1 \mathbf{a}_1),$$
$$d_{12} = v/v_{12},$$

where

$$Q_1 = \big(h(x_{1i}, x_{1j})\big), \qquad i, j = 1, 2, \dots, m,$$

$$Q_2 = \big(h(x_{2i}, x_{2j})\big), \qquad i, j = 1, 2, \dots, n,$$

$$h(x, y) = \min(x, y) - (x + y) + \tfrac{1}{2}(x^2 + y^2) + \tfrac{1}{3}.$$

PROOF. For $j = 2, 3, \dots, m$ let $\mathbf{a}_j$ be a $m \times 1$ vector of zeroes with 1 in the jth place and -1 in the $(j-1)$st place. For $j = 2, 3, \dots, n$ define $n \times 1$ vectors $\mathbf{b}_j$ similarly. Define

$$V_{ij} = (\mathbf{a}_i \times b_j)'\mathbf{F}.$$

For the prior as specified in Section 2 it can be checked that

(i) $\mathrm{Cov}[V_{ij}, V_{rs}] = 0$ if $i \neq r$ or $j \neq s$;
(ii) $\mathrm{Var}[V_{11}] > v_0$;
(iii) $\mathrm{Var}[V_{i1}] = (x_{1i} - x_{1i-1})(v_1 + v_{12}\mathbf{b}_1'Q_2\mathbf{b}_1)$, $i \geq 2$;
(iv) $\mathrm{Var}[V_{ij}] = (x_{2j} - x_{2j-1})(v_2 + v_{12}\mathbf{a}_1'Q_1\mathbf{a}_1)$, $j \geq 2$;
(v) $\mathrm{Var}[V_{ij}] = (x_{1i} - x_{1i-1})(x_{2j} - x_{2j-1})v_{12}$, $i, j \geq 2$.

Using these facts we may rewrite the log posterior density and the result follows upon noting that

$$V_{i1} = \alpha(F, i) - \alpha(F, i-1),$$

$$V_{1j} = \beta(F, j) - \beta(F, j-1),$$

$$V_{ij} = F_{ij} + F_{i-1\,j-1} - F_{i-1j} - F_{ij-1}. \qquad \square$$

The following corollary is easily shown.

COROLLARY 4.2. *For the special case where*

$$x_{1i} = (2i-1)/2m, \qquad x_{2j} = (2j-1)/2n,$$

the estimate $\hat{\mathbf{F}}$ *is chosen to minimise*

$$(\mathbf{y} - \mathbf{F})'(\mathbf{y} - \mathbf{F}) + d_1 \sum_{i=2}^{m} \frac{(\bar{F}_{i.} - \bar{F}_{i-1.})^2}{x_{1i} - x_{1i-1}}$$

$$+ d_2 \sum_{j=2}^{n} \frac{(\bar{F}_{.j} - \bar{F}_{.j-1})^2}{x_{2j} - x_{2j-1}} + d_{12} \sum_{i=2}^{m} \sum_{j=2}^{n} \frac{(F_{ij} + F_{i-1\,j-1} - F_{i-1j} - F_{ij-1})^2}{(x_{1i} - x_{1i-1})(x_{2j} - x_{2j-1})}.$$

This demonstrates how the two-dimensional technique may be regarded as a form of analysis of variance incorporating smoothness assumptions on the underlying regression process or alternatively as generalised variance components analysis.

5. Asymptotics: the general case. Define

$$R(\hat{F}, F) = \frac{1}{N}\sum_{\iota=1}^{N}(\hat{F}(\mathbf{x}_\iota) - F(x_\iota))^2.$$

Taking expectation with respect to the error distribution gives

$$ER(\hat{F}, F) = \frac{1}{N}\mathbf{F}'(I - A)^2\mathbf{F} + \frac{v}{N}\operatorname{trace}(A^2),$$

where $\mathbf{F} = (F(\mathbf{x}_1), F(\mathbf{x}_2), \dots, F(\mathbf{x}_N))'$ and A is the $N \times N$ matrix such that

$$\hat{\mathbf{F}} = A\mathbf{y}.$$

THEOREM 5.1. *For $F \in H$ we have that*

$$\begin{aligned} ER(\hat{F}, F) \le{}& \frac{c_1}{N}\int_0^1\left[\int_0^1 F_1(x_1, v)\,dv\right]^2 dx_1 \\ &+ \frac{c_2}{N}\int_0^1\left[\int_0^1 F_2(u, x_2)\,du\right]^2 dx_2 + \frac{c_{12}}{N}\int_0^1\int_0^1 F_{12}(u, v)^2\,du\,dv \\ &+ \left(\frac{v}{4}\right)\left[\frac{12}{N} + \frac{1}{c_1} + \frac{1}{c_2} + \frac{1}{c_{12}}\right]. \end{aligned}$$

PROOF. By Lemma 4.1 of Craven and Wahba (1979)

$$\begin{aligned} \mathbf{F}'(I - A)^2 F \le{}& c_1\int_0^1\left[\int_0^1 F_1(x_1, v)\,dv\right]^2 dx_1 \\ &+ c_2\int_0^1\left[\int_0^1 F_2(u, x_2)\,du\right]^2 dx_2 + c_{12}\int_0^1\int_0^1 F_{12}^2(u, v)\,du\,dv. \end{aligned}$$

Hence we need only consider bounding trace(A^2). From the remarks after Theorem 3.1 we have that

$$A = A_0 + E,$$

where

$$A_0 = QM^{-1}, \qquad E = vM^{-1}11'M^{-1}/1'M^{-1}1$$

with

$$Q = \left(Q(\mathbf{x}_\iota, \mathbf{x}_j)\right), \qquad M = Q + vI,$$

for

$$Q(\mathbf{x}, \mathbf{y}) = v_1 h(x_1, y_1) + v_2 h(x_2, y_2) + v_{12} h(x_1, y_1) h(x_2, y_2),$$

where

$$h(x, y) = \min(x, y) - (x + y) + \tfrac{1}{2}(x^2 + y^2) + \tfrac{1}{3}.$$

For matrices A and B we write $A \le B$ to mean that the eigenvalues of A are smaller than the corresponding eigenvalues of B. Clearly $0 \le A_0 \le I$. We have that

$$\begin{aligned} I - A &= vM^{-1} - E \\ &= vM^{-1/2}\{I - M^{-1/2}11'M^{-1/2}/1'M^{-1}1\}M^{-1/2}. \end{aligned}$$

Since $M^{-1/2}11'M^{-1/2} \ge 0$ and $\text{trace}[M^{-1/2}11'M^{-1/2}/1'M^{-1}1] = 1$ we have

$$\begin{aligned} &0 \le M^{-1/2}11'M^{-1/2}/1'M^{-1}1 \le I \\ &\Rightarrow I - A \ge 0 \Rightarrow A \le I \Rightarrow E \le I. \end{aligned}$$

Hence

$$\begin{aligned} \text{trace}(A^2) &= \text{tr}(A_0^2) + 2\,\text{tr}(A_0 E) + \text{tr}(E^2) \\ &\le \text{tr}(A_0^2) + 3 \end{aligned}$$

since $\text{rank}(A_0 E) \le \text{rank}(E) = 1$. Also

$$\begin{aligned} \text{trace}(A_0^2) &= \text{tr}[QM^{-1}]^2 \\ &\le \text{tr}\frac{[Q]}{4v} \quad \text{as} \left(\frac{x}{x+v}\right)^2 \le \frac{x}{4v} \\ &\le \frac{1}{4v}[Nv_1 + Nv_2 + Nv_{12}] \end{aligned}$$

as $Q(\mathbf{x}, \mathbf{y}) \le v_1 + v_2 + v_{12}$. Hence the result. □

Hence $ER(\hat{\mathbf{F}}; F) \to 0$ provided

(i) $c_1 \to \infty,\ c_2 \to \infty,\ c_{12} \to \infty$.
(ii) $c_1/N \to 0,\ c_2/N \to 0,\ c_{12}/N \to 0$.

Expressed in terms of the variances of the prior these requirements become

(i)′ $v_1 \to 0,\ v_2 \to 0,\ v_{12} \to 0$.
(ii)′ $Nv_1 \to \infty,\ Nv_2 \to \infty,\ Nv_{12} \to \infty$.

Hence as N increases the prior must be tighter [due to (i)′], but not too tight [due to (ii)′].

6. Asymptotics: the grid case. In this section we bound $ER(\hat{F}, F)$ for data on a grid in two dimensions

THEOREM 6.1. *Suppose we have $N = mn$ observations on a grid*

$$\{(x_{1i}, x_{2j}) : 1 \le i \le m, 1 \le j \le n\}.$$

Thus, for $F \in H$

$$ER(\hat{F}, F) \le \frac{1}{N}P(F) + O\left[\left(\frac{B_1(v_1 + v_{12})}{n}\right)^{1/2}\right] + O\left[\left(\frac{B_2(v_2 + v_{12})}{m}\right)^{1/2}\right] + O\left[(B_1 B_{12} v_{12})^{1/2}\log(B_1 B_2 v_{12})\right],$$

where

$$B_1 = \max(x_{1i} - x_{1i-1}), \qquad B_2 = \max(x_{2j} - x_{2j-1}),$$

and $P(F)$ is as in Section 3.

PROOF. From Theorem 4.1 we have that

$$\hat{\mathbf{F}} = A\mathbf{y},$$

where

$$A = \left[I + d_1(G_1 \times \mathbf{b}_1\mathbf{b}_1') + d_2(\mathbf{a}_1\mathbf{a}_1' \times G_2) + d_{12}(G_1 \times G_2)\right]^{-1},$$

where $G_1 = (g_{ij})$ is an $m \times m$ symmetric tri-diagonal matrix with

$$g_{ii} = (x_{1i+1} - x_{1i})^{-1} + (x_{1i} - x_{1i-1})^{-1}, \qquad 2 \le i \le m - 1,$$

$$g_{11} = (x_{12} - x_{11})^{-1},$$

$$g_{mm} = (x_{1m} - x_{1m-1})^{-1},$$

$$g_{ii-1} = -(x_{1i} - x_{1i-1})^{-1}, \qquad 2 \le i \le m.$$

G_2 is a similarly defined $n \times n$ symmetric tri-diagonal matrix, $\mathbf{a}_1, \mathbf{b}_1, d_1, d_2, d_{12}$ are as in Theorem 4.1 and $A \times B$ denotes the Kronecker product of A and B [see Bellman (1970)]. As in Section 5

$$ER(\hat{F}, F) = \frac{1}{N}\mathbf{F}'(I - A)^2\mathbf{F} + \frac{v}{N}\operatorname{tr}(A^2)$$

and

$$\mathbf{F}'(I - A)^2\mathbf{F} \le P(F).$$

We proceed to bound $\operatorname{tr}(A^2)$.

The following lemma will be proved later.

LEMMA 6.2

$$A \le \left[I + \delta_1(H_1 \times J_2) + \delta_2(J_1 \times H_2) + \delta_{12}(H_1 \times H_2)\right]^{-1},$$

where

$$\delta_1 = d_1/n^2 B_1, \qquad \delta_2 = d_2/m^2 B_2, \qquad \delta_{12} = d_{12}/B_1 B_2.$$

$J_1(J_2)$ is an $m \times m(n \times n)$ matrix of ones. $H_1(H_2)$ is an $m \times m(n \times n)$ matrix

with

$$h_{ij} = \begin{cases} 2, & i = j, \quad 1 < i < m, \\ 1, & i = j, \quad i = 1, m, \\ -1, & i = j + 1 \ \text{ or } \ j = i + 1, \\ 0, & \text{otherwise}. \end{cases}$$

Noting that

(i) $H_1 J_1 = J_1 H_1$, $H_2 J_2 = J_2 H_2$,
(ii) the first eigenvalue of $J_1(J_2)$ is $m(n)$ and the rest zeroes,
(iii) the eigenvalues of H_1 are

$$2\left(1 - \cos\left(\frac{\pi r}{m}\right)\right), \qquad r = 0, 1, \dots, m-1,$$

and similarly for H_2, we have that

$$\operatorname{tr}(A^2) \le 1 + S_1 + S_2 + S_{12},$$

where

$$S_1 = \sum_{r=1}^{m-1} \frac{1}{\left[1 + 2n\delta_1\left(1 - \cos\left(\frac{\pi r}{m}\right)\right)\right]^2},$$

$$S_2 = \sum_{r=1}^{n-1} \frac{1}{\left[1 + 2m\delta_2\left(1 - \cos\left(\frac{\pi r}{n}\right)\right)\right]^2},$$

and

$$S_{12} = \sum_{r=1}^{m-1} \sum_{s=1}^{n-1} \frac{1}{\left[1 + 4\delta_{12}\left(1 - \cos\left(\frac{\pi r}{m}\right)\right)\left(1 - \cos\left(\frac{\pi s}{n}\right)\right)\right]^2}.$$

The terms S_1, S_2, and S_{12} can be bounded as follows:
Choose $M > 0$ such that

$$Mx^2 \le 2(1 - \cos(x)) \text{ for } x \text{ in } [0, \pi].$$

Then

$$\begin{aligned} S_1 &\le \sum_{r=1}^{m-1} \frac{1}{\left[1 + n\delta_1 M\left(\frac{\pi r}{m}\right)^2\right]^2} \\ &\le \frac{m}{\pi} \int_0^{\Pi} \frac{dx}{\left[1 + n\delta_1 M x^2\right]^2} \\ &= O\left(\frac{m}{(n\delta_1)^{1/2}}\right) = O\left[mn\left(\frac{B_1}{nd_1}\right)^{1/2}\right]. \end{aligned}$$

Similarly,

$$S_2 = O\left(\frac{n}{(m\delta_2)^{1/2}}\right) = O\left[mn\left(\frac{B_2}{md_2}\right)^{1/2}\right],$$

$$S_{12} \le \sum_{r=1}^{m-1}\sum_{s=1}^{n-1} \frac{1}{\left[1 + \delta_{12}M^2\left(\frac{\pi r}{m}\right)^2\left(\frac{\pi s}{n}\right)^2\right]^2}$$

$$\le \frac{mn}{\pi^2}\int_0^{\pi}\int_0^{\pi}\frac{dx\,dy}{\left(1 + \delta_{12}M^2x^2y^2\right)^2}$$

$$= O\left[mn \log\frac{(\delta_{12})}{\sqrt{\delta_{12}}}\right]$$

$$= O\left[mn\left(\frac{B_1B_2}{d_{12}}\right)^{1/2}\log\left(\frac{B_1B_2}{d_{12}}\right)\right].$$

Since all the elements of $\mathbf{a}_1$ are positive and the largest element of Q_1 is less than or equal to one we have

$$\mathbf{a}_1'Q_1\mathbf{a}_1 \le \mathbf{a}_1'J_1\mathbf{a}_1$$
$$= 1.$$

Hence $1/d_1 = O[v_1 + v_{12}]$. Similarly, $1/d_2 = O[v_2 + v_{12}]$. Hence the theorem follows. □

Proof of Lemma 6.2. (i) We show

$$\mathbf{a}_1\mathbf{a}_1' \ge J_1/m^2.$$

Both $\mathbf{a}_1\mathbf{a}_1'$ and J_1 have only one nonzero eigenvalue and

$$\operatorname{tr}(\mathbf{a}_1\mathbf{a}_1') = \operatorname{tr}(\mathbf{a}_1'\mathbf{a}_1)$$
$$\ge \frac{1}{m}\left(\sum a_{1i}\right)^2$$
$$= \frac{1}{m}$$
$$= \frac{1}{m^2}\operatorname{tr}(J_1).$$

Similarly, $\mathbf{b}_1\mathbf{b}_1' \ge J_2/n^2$.

(ii) $G_1 \ge (1/B_1)H_1$. Let $d_i = (x_{1i} - x_{1i-1})^{-1}$, $i = 2, 3, \dots, m$. Then, for any vector $\mathbf{s} = (s_1, s_2, \dots, s_m)'$,

$$\mathbf{s}'G_1\mathbf{s} = d_2(s_2 - s_1)^2 + d_3(s_3 - s_2)^2 + \cdots + d_m(s_m - s_{m-1})^2$$
$$\ge \left[(s_2 - s_1)^2 + (s_3 - s_2)^2 + \cdots + (s_m - s_{m-1})^2\right]/B_1$$
$$= \mathbf{s}'H_1\mathbf{s}/B_1.$$

Similarly, $G_2 \geq (1/B_2)H_2$. Hence the lemma follows. □

COROLLARY 6.3. *If* $B_1 = O(1/m)$ *and* $B_2 = O(1/n)$, *then for*

$$v_1 = R_1 N^{-1/3}, \qquad v_2 = R_2 N^{-1/3} \text{ and } v_{12} = R_{12} N^{-1/3}(\log N)^{-2/3},$$

where R_1, R_2, *and* R_{12} *are constants, we have*

$$ER(\hat{F}, F) \leq O\left[N^{-2/3}(\log N)^{2/3}\right].$$

The above bound is much tighter than that obtained in Section 5.

7. Wahba's technique. Wahba (1979) has considered the use of "thin plate" splines to smooth surfaces in higher dimensions. In two dimensions the simplest form of such splines involves choosing $\hat{F}$ to minimise

$$\sum_{i=1}^{n}\left(y_i - F(x_{1i}, x_{2i})\right)^2 + c\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{11}^2 + 2F_{12}^2 + F_{22}^2, \tag{7.1}$$

where $F_{ij} = \partial^2 F/\partial x_i\, \partial x_j$.

Wahba (1979) states that the solution $\hat{F}_c$ has a representation

$$\hat{F}_c(\mathbf{x}) = d_0 + d_1 x_1 + d_2 x_2 + \sum_{j=1}^{n} b_j E(\mathbf{x}, \mathbf{x}_j),$$

where

$$E(\mathbf{x}, \mathbf{y}) = \frac{1}{8\pi}|\mathbf{x} - \mathbf{y}|^2 \log|\mathbf{x} - \mathbf{y}|$$

with

$$|\mathbf{x} - \mathbf{y}|^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2.$$

It is shown in Wahba and Wendelberger (1980) that the Bayesian procedure corresponding to this technique is to assume that

$$y_i = F(\mathbf{x}_i) + e_i, \qquad i = 1, 2, \ldots, n,$$

with $\{e_i\}$ iid $N(0, v)$ and to put a prior on F which is the same as the stochastic process

$$\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + v_1^{1/2} Z(\mathbf{x}),$$

where

$$\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2) \sim N(0, v_0 I), \qquad v_0 \to \infty,$$
$$v_1 = v/c,$$

and $Z(\mathbf{x})$ is a mean zero Gaussian process with covariance function

$$Q(\mathbf{x}, \mathbf{y}) = E(\mathbf{x}, \mathbf{y}) - \sum_{j=1}^{3} P_j(\mathbf{x}) E(S_j, \mathbf{y}) - \sum_{j=1}^{3} P_j(\mathbf{y}) E(S_j, \mathbf{x}) + \sum_{i=1}^{3}\sum_{j=1}^{3} P_i(x) P_j(y) E(S_i, S_j),$$

where S_1, S_2, S_3 are chosen so that

$$\sum_{j=1}^{3} \alpha_j \phi_j(S_k) = 0 \quad \text{for } k = 1,2,3,$$

$$\Rightarrow \alpha_j = 0 \quad \text{for } j = 1,2,3$$

for any basis $\{\phi_1, \phi_2, \phi_3\}$ of the space of polynomials of total degree 1 or 0 and P_j is such that $P_j(S_k) = \delta_{jk}$.

Two criticisms of this technique can be made. First, only one smoothing parameter is used, suggesting that the function to be estimated is equally smooth in all directions. This is rarely true. Second, the equivalent Bayesian formulation involves the use of a complicated covariance function seemingly unrelated to that used in 1-d. Wahba (1981) suggests a possible answer to the first criticism. In 2-d her suggestion corresponds to a roughness penalty of the form

$$c\int_{R^2} F^2_{x_1x_1} + 2\theta F^2_{x_1x_2} + \theta^2 F^2_{x_2x_2}. \tag{7.2}$$

The technique of 2 was originally proposed to address the second criticism but has resulted in a technique which also overcomes (at least theoretically) the first criticism.

The roughness penalty (7.1) is invariant under rotation of the x_1 and x_2 axes; the penalty (7.2) is not. Likewise the Bayesian technique of this paper is not invariant under rotation of the axes.

The roughness penalty (7.1) is the closest to the Bayesian technique in the sense that the infinitely smoothed estimate is the least-squares plane while for the Bayesian technique the infinitely smoothed estimate is a constant. Wahba (1979) considers roughness penalties involving higher derivatives leading to higher-order "thin plate" splines. The roughness penalty involving derivatives of order 3 is

$$C\int_{R^2} F^2_{111} + F^2_{222} + 3F^2_{112} + 3F^2_{122}, \tag{7.3}$$

where

$$F_{ijk} = \frac{\partial^3 F}{\partial x_i\, \partial x_j\, \partial x_k}.$$

Barry (1983) describes models for incorporating stronger smoothness assumptions in the Bayesian framework of this paper.

8. Simulation study. A simulation study was carried out to compare Wahba's techniques using roughness penalties (7.1) and (7.3) with the Bayesian technique of Section 2. For each of four underlying regression functions data were generated on the grid

$$(x_{1i}, x_{2j}): \{1 \le i \le 10, 1 \le j \le 10\},$$

where

$$x_{1i} = (2i - 1)/20, \qquad x_{2j} = (2j - 1)/20,$$

by setting

$$y_{ij} = F(x_{1i}, x_{2j}) + e_{ij},$$

where $\{e_{ij}\}$ are iid $N(0, v)$. The four underlying regression functions used were:

F_1: $6144(xy)^5(1 - xy)^7$;
F_2: $1.5\sin(12x)\sin(12y)$;
F_3: $(L(x) + 3L(y) + L(x)L(y))/8$,
where

$$\begin{aligned} L(x) &= 8x, & 0 \le x \le 0.25, \\ &= 2 - 8(x - 0.25), & 0.25 \le x \le 0.5, \\ &= 6(x - 0.5), & 0.5 \le x \le 0.75, \\ &= 1.5 - 6(x - 0.75), & 0.75 \le x \le 1.0; \end{aligned}$$

F_4: $1.5Z_1(x)Z_2(y)$, where Z_1 and Z_2 are independent Brownian motions.

F_1, F_2, and F_3 each have maximum value 1.5. F_1 is slowly changing and infinitely differentiable; F_2 changes quickly, but is also infinitely differentiable; F_3 is continuous, but only piecewise differentiable, while F_4 is a sample path from a stochastic process which has continuous, nowhere differentiable sample paths. Three values for v were used: $v = 0.01$, 0.0625, and 0.25 (corresponding to standard deviations of 0.1, 0.25, and 0.5, respectively).

For each combination of regression function and error variance v, 50 repetitions were carried out and the average mean squared error obtained using the three techniques is recorded in Table 1. In all cases the smoothing parameters

TABLE 1
Average mean squared residual using (i) Bayesian technique, (ii) Wahba [roughness penalty (7.1)], and (iii) Wahba [roughness penalty (7.3)]

function	$v = 0.01$	$v = 0.0625$	$v = 0.25$
F_1	0.00654	0.0273	0.0725
	0.00575	0.0195	0.0526
	0.00390	0.0175	0.0553
F_2	0.00886	0.0541	0.1629
	0.00983	0.0627	0.1791
	0.00914	0.0429	0.1252
F_3	0.00336	0.0159	0.0495
	0.00775	0.0239	0.0559
	0.00605	0.0230	0.0568
F_4	0.01034	0.0626	0.2140
	0.01034	0.0648	0.2284
	0.01034	0.0775	0.2865

were chosen by generalised cross validation as described in Craven and Wahba (1979).

The comparison is not clearcut. A general observation would be that the Bayesian technique is best for rougher functions. For the smoothest function F_1 Wahba's techniques do far better than the Bayesian technique. However, as the roughness of the underlying regression function increases the Bayesian technique becomes more competitive and does best for F_3 and F_4.

The comparison is a little clouded for $v = 0.01$. Here the Bayesian technique does best for F_2 and all techniques are equivalent for F_4—in fact they all opt to do no smoothing at all.

The design of a simulation study to compare different smoothing techniques is difficult since the choice of test functions seems crucial. The above study, using four quite different functions, suggests that the Bayesian technique works well when the underlying regression function has limited smoothness properties. However, the decision as to which technique to use in a particular situation seems very difficult and needs further study.

9. Regression in higher dimensions. The two-dimensional technique was based on a decomposition of the regression function F into four parts analogous to a decomposition widely used in the parameterisation of two-way analysis of variance: overall mean, row effects, column effects, and interaction terms. Continuing the analogy into higher dimensions leads in a straightforward manner to the appropriate generalisation of the two-dimensional case.

In three dimensions, for example, we use the decomposition

$$F(x_1, x_2, x_2) = \mu + \alpha_1(x_1) + \alpha_2(x_2) + \alpha_3(x_3) + \alpha_{12}(x_1, x_2) + \alpha_{23}(x_2, x_3) + \alpha_{13}(x_1, x_3) + \alpha_{123}(x_1, x_2, x_3),$$

where

$$\mu = \int\int\int F(u, v, w),$$

$$\alpha_1(x_1) = \int\int F(x_1, v, w) - \mu$$

$[\alpha_2, \alpha_3$ similarly],

$$\alpha_{12}(x_1, x_2) = \int F(x_1, x_2, w) - \int\int F(x_1, v, w) - \int\int F(u, x_2, w) + \int\int\int F(u, v, w)$$

$[\alpha_{23}, \alpha_{13}$ similarly], and

$$\alpha_{123}(x_1, x_2, x_3) = F(x_1, x_2, x_3) - \int F(x_1, x_2, w) - \int F(u, x_2, x_3) - \int F(x_1, v, x_3) + \int\int F(x_1, v, w) + \int\int F(u, x_2, w) + \int\int F(u, v, x_3) - \int\int F(u, v, w),$$

where all integrals are from 0 to 1. Appropriately adjusted Brownian sheet priors can be placed on each term and Bayes theorem applied to get $\hat{F}$.

The extension to higher dimensions is described in detail in Barry (1983) where consistency results are also demonstrated. As the number of dimensions increases the number of prior parameters increases and it may be necessary to include higher-order interaction terms in the error term as is often done in multi-way ANOVA.

Acknowledgments. This research was carried out at Yale University as part of the requirements for a Ph.D. degree in statistics. I would like to thank Professor John Hartigan for his help and encouragement during the course of this research. Thanks are also due to Professor Grace Wahba for sending me copies of her work in this area.

REFERENCES

BARRY, D. (1983). Nonparametric Bayesian regression. Ph.D. thesis, Yale Univ.

BELLMAN, R. (1970). *Introduction to Matrix Analysis.* McGraw-Hill, New York.

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.

KIMELDORF, G. and WAHBA, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82–95.

SCHOENBERG, I. J. (1964). Spline functions and the problem of graduation. *Proc. Nat. Acad. Sci. U.S.A.* **52** 947–950.

SHEPP, L. (1966). Radon–Nikodym derivatives of Gaussian measures. *Ann. Math. Statist.* **37** 321–354.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

WAHBA, G. (1979). Convergence rates of "thin plate" smoothing splines when the data are noisy. Technical Report 557, Dept. Statistics, Univ. of Wisconsin, Madison.

WAHBA, G. (1981). Data-based optimal smoothing of orthogonal series density estimates. *Ann. Statist.* **9** 146–156.

WAHBA, G. and WENDELBERGER, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review* **108** 1122–1143.

ZIMMERMAN, G. J. (1972). Some sample-function properties of the two-parameter Gaussian process. *Ann. Math. Statist.* **43** 1235–1246.

DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE
CORK
IRELAND