# SURFACE FITTING WITH SCATTERED NOISY DATA ON EUCLIDEAN D-SPACE AND ON THE SPHERE

GRACE WAHBA

ABSTRACT. An overview of cross validated spline methods for smoothing noisy data in the plane, in Euclidean $d$-space, and on the sphere is given. Cross-validated thin plate smoothing splines are reviewed and an efficient numerical algorithm for computing them for problems with up to several hundred data points is described. Some numerical results for a two-dimensional example are given. A theory of vector splines for smoothing noisy vector data on the sphere is given. The use of generalized cross-validation to estimate both the smoothing parameter as well as the relative energy to be assigned to the divergent and nondivergent part of the smoothed vector field is described and tested numerically on simulated upper air horizontal wind fields.

**1. Introduction; an overview of cross-validated smoothing splines.** It is assumed that data $y = (y_1, \ldots, y_n)'$ arise according to the model

$$y_i = f(P_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where $P_i \in S$, some index set (e.g., Euclidean $d$-space, the sphere, torus, etc.). The function $f$ is assumed to be a smooth function in some reproducing kernel (r.k.) Hilbert space $X$ of real-valued or vector-valued functions on $S$. The $\{\varepsilon_i\}$ are independent zero mean measurement errors with common unknown variance, and it is desired to recover an estimate of $f$ given $y = (y_1, \ldots, y_n)'$. The estimate $f_\lambda$ of $f$ will be taken as the minimizer in $X$ of

$$(1.1) \qquad \frac{1}{n} \sum_{i=1}^{n} (y_i - f(P_i))^2 + \lambda J(f),$$

where $J(f)$ is a seminorm on $X$ with $M$-dimensional null space spanned by $\phi_1, \ldots, \phi_M$, $M < n$. Here $J^{1/2}(f)$ can be taken as the norm of the orthogonal projection of $f$ onto $X_1$ where $X_1$ is the orthocomplement of the span of the $\{\phi_\nu\}$ in $X$. If the $n \times M$ matrix $T$ with $(i, \nu)$-th entry

$\phi_\nu(P_i)$ is of rank $M$, then the minimizer of (1.1) exists uniquely and always has a representation of the form

$$(1.2) \qquad f_\lambda(P) = \sum_{i=1}^{n} c_i \xi_i(P) + \sum_{\nu=1}^{M} d_\nu \phi_\nu(P),$$

$P \in S$, where the $\xi_i$ can be determined from an r.k. or semikernel for $X$ with seminorm $J^{1/2}$, and $c = (c_1, \ldots, c_n)'$ and $d = (d_1, \ldots, d_M)'$ satisfy a system of $n + M$ linear equations. For more details, see [15], [24], §2 below and references cited there.

The most famous special case of $J$ is surely

$$(1.3) \qquad J(f) = \int_0^1 (f^{(m)}(x))^2 dx,$$

which for $m = 2$ leads to the celebrated cubic smoothing splines (which are generally not computed using (1.2) due to the existence of a local support basis for the span of the $\{\xi_i\}$ in the special case of (1.3)). In this paper we will primarily be considering generalizations of $J$ of (1.3) to Euclidean $d$-space of the form (for $d = 2$)

$$(1.4) \qquad J(f) = \iint_{-\infty}^{\infty} \sum_{\nu=0}^{m} \left( \frac{\partial^m f}{\partial x_1^\nu \partial x_2^{m-\nu}} \right)^2 dx_1 dx_2,$$

which lead to the thin plate splines, and to the sphere of the form (for $m$ even)

$$(1.5) \qquad J(f) = \int_S (\Delta^{m/2} f)^2 dP$$

where $\Delta$ is the Laplacian on the sphere.

The parameter $\lambda$ controls the tradeoff between the "roughness" of the solution as measured by $J(f)$, and the infidelity of the solution to the data, as measured by $(1/n) \sum_{i=1}^{n} (y_i - f(P_i))^2$. The visual appearance of $f_\lambda$ can be quite sensitive to the choice of $\lambda$, and estimates of derivatives of $f$ obtained by differentiating $f_\lambda$ even more sensitive. The value $\lambda^*$ of $\lambda$ which minimizes the predictive mean square error defined by $\sum_{i=1}^{n}(f_\lambda(P_i) - f(P_i))^2$ can be estimated from the data by the method of generalized cross-validation (GCV). The GCV estimate $\hat{\lambda}$ of $\lambda$ is the minimizer of the cross-validation function $V(\lambda)$ given by

$$(1.6) \qquad V(\lambda) = \frac{\frac{1}{n} \|(I - A(\lambda))y\|^2}{\left( \frac{1}{n} \mathrm{Tr}(I - A(\lambda)) \right)^2},$$

where $A(\lambda)$ is the so called "influence matrix", which satisfies

(1.7)
$$\begin{bmatrix} f_\lambda(P_1) \\ \vdots \\ f_\lambda(P_n) \end{bmatrix} = A(\lambda)y.$$

For $n$ moderate to large, $\hat{\lambda}$ so obtained is known to be a good estimate of $\lambda^*$. For a discussion of the method and its properties, see [23], [20], [2], [11] and [29].

In general, to find $f_\lambda$ for given $\lambda$ requires the solution of a linear system of the order of the number, $n$, of data points, and to find the minimizer of $V(\lambda)$ entails the solution of an eigenvector-eigenvalue problem of size $(n - M) \times (n - M)$ (see §2 below). For $n$ very large then, it has been proposed [25, 26] that (1.1) be minimized in the span $X^{(p)} \subset X$ of $p$ suitably chosen basis functions, where $p$ is chosen large enough so that $\lambda$ (and not $p$) is the smoothing parameter. More specifically one wants the minimizer of (1.1) in $X^{(p)}$ to be a good approximation to the minimizer of (1.1) in $X$ if $f$ is a "smooth" function. Natural choices of these basis functions are sines and cosines if $X$ is a space of periodic functions on the circle, and $B$-splines [3] of degree $2m-1$ if $J(f) = \int_0^1 (f^{(m)}(x))^2 dx$. Basis functions which are a generalization of $B$-splines to Euclidean $d$-space ($E^d$) have been suggested in [25], see also [7].

The cross-validation function $V(\lambda)$ is still defined by (1.6) and (1.7), but now $A(\lambda)$ will depend on the basis functions used. In this case an $(n - M) \times p$ singular value decomposition can be used to determine $\hat{\lambda}$ instead of an $(n - M) \times (n - M)$ eigenvalue-eigenvector decomposition, see §3 below. Bates and Wahba [1] give efficient methods for computing $\hat{\lambda}$ and $f_{\hat{\lambda}}$ when $n$ is very large and $M < p < n$ basis functions are used, including an efficient truncation procedure for the singular value decomposition.

In §2 we describe an efficient algorithm from J. Wendelberger's thesis [32] for computing cross-validated thin plate splines in $d$ dimensions, and show some numerical results. In §3 we extend these ideas to vector observations on the sphere and investigate the ability of the GCV method to govern the relative energy assigned to the divergent and nondivergent part of the estimated vector field. The estimates appear to be very good from a mean square error point of view.

A good value of the parameter $m$ governing the number of derivatives in $J$ can also be estimated by GCV as can $d - 1$ relative scale factors in $E^d$, see [10], [31] and [32]. It is believed that, if $n$ is large, $\lambda$ as well as several parameters in $J$ can be estimated by GCV provided the several parameters are chosen so that distinct values of $\lambda$ and the several parameters are always associated with r.k.'s which correspond to perpendicular stochastic processes, see [28] for a brief discussion of this point, which we do not pursue further here.

**2. Thin plate splines on $E^d$.** In the thin plate spline case $J(f) = J_m^d(f)$ is given by

$$J_m^d(f) = \sum_{i_1=1,\ldots,i_m=1}^{i_1=d,\ldots,i_m=d} \int \cdots \int \left(\frac{\partial^m f}{\partial x_{i_1} \cdots \partial x_{i_m}}\right)^2 dx_1 \cdots dx_d.$$

In particular, for $m = 2$, $d = 2$,

$$J_2^2(f) = \iint \left[\left(\frac{\partial^2 f}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2}\right)^2\right] dx_1 dx_2.$$

The space $X$ is the space of all generalized functions for which all partial derivatives of order less or equal to $m$ exist in $L_2(E^d)$. For $X$ to be a reproducing kernel Hilbert space (that is, for the evaluation functionals in $X$ to be bounded with respect to $J$), it is necessary that $2m - d > 0$. The null space of $J_m^d$ is the span of the $M = \binom{d+m-1}{d}$ polynomials in $E^d$ of total degree less than $m$. Duchon [5, 6] has given a semikernel from which the explicit representation of $f$ given in Theorem 1 below has been obtained. (See also [16] and [31].)

Before giving a formula for $f_\lambda$ we need some notation. Let $t \in E^d$, $t = (x_1, x_2, \ldots, x_d)$. If $s = (y_1, \ldots, y_d)$, then $|t - s| = (\sum_{i=1}^d (x_i - y_i)^2)^{1/2}$. Let $E(\tau)$ be defined by

$$E(\tau) = \begin{cases} \theta_m^d \tau^{2m-d} \log \tau & d \text{ even} \\ \theta_m^d \tau^{2m-d} & d \text{ odd} \end{cases}$$

where

$$\theta_m^d = \begin{cases} \dfrac{(-1)^{d/2+1}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} & d \text{ even} \\[3mm] \dfrac{(-1)^m \Gamma(d/2 - m)}{2^{2m}\pi^{d/2}(m-1)!} & d \text{ odd.} \end{cases}$$

Let $E_{t_i}(t)$ be the function defined by

$$(2.1) \qquad\qquad E_{t_i}(t) = E(|t - t_i|),$$

and let $\phi_1, \ldots, \phi_M$ be the $M$ polynomials of total degree less than $m$; for example, if $d = 2$, $m = 2$, then $M = 3$, $\phi_1(x_1, x_2) = 1$, $\phi_2(x_1, x_2) = x_1$, and $\phi_3(x_1, x_2) = x_2$.

THEOREM 1. *The minimizer $f_\lambda$ in $X$ of*

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(t_i))^2 + \lambda J_m^d(f)$$

*is given by*

$$(2.2) \qquad\qquad f_\lambda = \sum_{i=1}^n c_i E_t + \sum_{\nu=1}^M d_\nu \phi_\nu$$

*where* $c = (c_1, \ldots, c_n)'$ *and* $d = (d_1, \ldots, d_M)'$ *are determined by*

(2.3) $$(K + n\lambda I)c + Td = y$$

(2.4) $$T'c = 0$$

*where* $K$ *is the* $n \times n$ *matrix with* $(i, j)$-*th entry* $E(|t_i - t_j|)$ *and* $T$ *is the* $n \times M$ *matrix with* $(i, \nu)$-*th entry* $\phi_\nu(t_i)$.

If $Q$ is any $n \times (n - M)$ matrix satisfying $Q'Q = I_{n-M}$ and $Q'T = 0_{(n-M) \times M}$, then it can be shown [31] that

(2.5) $$I - A(\lambda) = n\lambda Q(Q'KQ + n\lambda I)^{-1}Q'$$

and

(2.6) $$n\lambda c = (I - A(\lambda))y.$$

An efficient numerical algorithm for obtaining $\hat{\lambda}$, $c$ and $d$ is given in Wendelberger's thesis [32]. It goes as follows.

1. Form the $QR$ decomposition of $T$ using LINPACK [4]

(2.7) $$T = (Q_1 : Q_2)\binom{R_1}{0}$$

where the matrix dimensions are $T$: $n \times M$; $Q_1$: $n \times M$; $Q_2$: $n \times (n - M)$; $R_1$: $M \times M$, $(Q_1 : Q_2)$ is orthogonal and $R_1$ is upper triangular. Matrix $Q$ in (2.5) will be taken as $Q_2$. Let $B = Q_2'KQ_2$.

2. Find the eigenvalue-eigenvector decomposition of $B \equiv Q_2'KQ_2 = UD_BU'$, where $U$ is orthogonal and $D$ is diagonal with diagonal entries $b_\nu$, $\nu = 1, 2, \ldots, n - M$. Then

(2.8) $$(I - A(\lambda))y = n\lambda Q_2U(D_B + n\lambda I)^{-1}U'Q_2'y$$

and

(2.9) $$\mathrm{Tr}(I - A(\lambda)) = \sum_{\nu=1}^{n-M} \frac{n\lambda}{b_\nu + n\lambda}.$$

Letting $w = (w_1, \ldots, w_{n-M}) = U'Q_2'y$ gives

(2.10) $$V(\lambda) = \frac{\frac{1}{n}\|(I - A(\lambda))y\|^2}{\left(\frac{1}{n}(\mathrm{Tr}(I - A(\lambda)))\right)^2} = \frac{\frac{1}{n}\sum_{\nu=1}^{n-M}\left(\frac{n\lambda}{b_\nu + n\lambda}\right)^2 w_\nu^2}{\left(\frac{1}{n}\sum_{\nu=1}^{n-M}\frac{n\lambda}{b_\nu + n\lambda}\right)^2}.$$

3. Find the minimizer $\hat{\lambda}$ of $V(\lambda)$ by global search in $\log_{10}\lambda$ on the right hand side of (2.10).

4. Obtain $c$ from

(2.11) $$c = Q_2U(D_B + n\hat{\lambda}I)^{-1}w.$$

Vector $d$ may be obtained by noting that $T = Q_1 R_1$, $Q_1' Q_1 = I_{M \times M}$ and premultiplying (2.3) by $Q_1'$ to obtain $Q_1' Kc + R_1 d = Q_1' y$.

5. Since $R_1$ is $M \times M$ upper triangular, $d$ is easily obtained by solving

$$(2.12) \qquad\qquad R_1 d = Q_1'(y - Kc).$$

A test function $f(x_1, x_2)$ was generated by Wendelberger and is shown in Figure 2.1. A regular $7 \times 7$ square array of 49 points $t_i$, $i = 1, 2, \ldots,$ 49, was selected with the middle point $(0, 0)$ and the spacing 1.0. Data $y_i$, $i = 1, 2, \ldots, 49$, was generated as $y_i = f(t_i) + \varepsilon_i$, $t_i = (x_{1i}, x_{2i})$, where the $\varepsilon_i$ were pseudorandom normally distributed random variables with zero mean and standard deviation .01. This standard deviation is about $1/8$ of the maximum height of $f$. Figure 2.2 shows $f_\lambda$ with $\lambda$ too big,
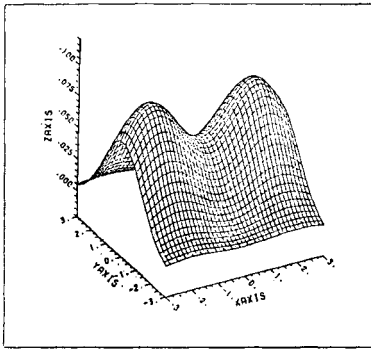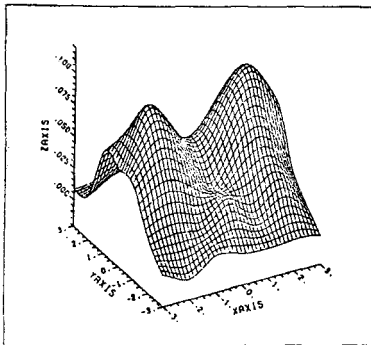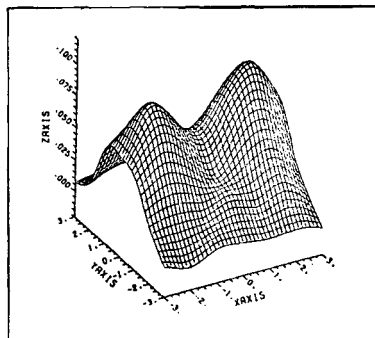


THE TEST FUNCTION.

FIGURE 2.1
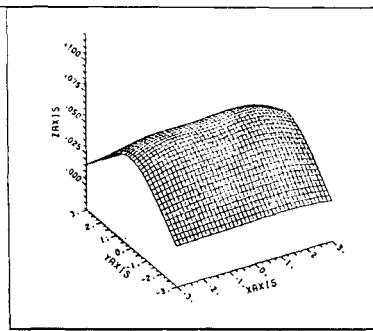


THE FITTED SURFACE $\lambda = 100 \hat{\lambda}$.

FIGURE 2.2



THE FITTED SURFACE $\lambda = .01 \hat{\lambda}$.

FIGURE 2.3



THE FITTED SURFACE $\lambda = \hat{\lambda}$.

FIGURE 2.4

Figure 2.3 shows $f_\lambda$ with $\lambda$ too small, and Figure 2.4 shows $f_{\hat\lambda}$. In this experiment the effectiveness of $\hat\lambda$ can be measured by the inefficiency

$$I = \frac{1}{n}\sum_{i=1}^{n}(f_{\hat\lambda}(P_i)) - f(P_i))^2 \Big/ \frac{1}{n}\sum_{i=1}^{n}(f_{\lambda^*}(P_i) - f(P_i))^2$$

which was 1.54. Theory as well as other numerical results show that $I \downarrow 1$ rapidly as $n$ becomes large, if $f$ is smooth. For further numerical results see [21].

**3. Vector splines on the sphere.** Wind fields, magnetic fields, etc., are measured on the surface of the earth discretely and with error. From this discrete data it is desired to estimate the field everywhere. In the case of wind fields, it is also desired to estimate the horizontal divergence $D$ and the vorticity $\zeta$ of the fields. We define vector splines on the sphere for this purpose, and show the results of a Monte Carlo study.

We will first describe univariate splines on the sphere and then use the results to define vector splines on the sphere. Let $P$ be a point on the sphere, $P = (\lambda, \phi)$, where $\lambda$ is its longitude ($0 \leq \lambda < 2\pi$) and $\phi$ is its latitude ($-\pi/2 \leq \phi \leq \pi/2$). (We are using $\lambda$ here for both longitude and the smoothing parameter. Which one is meant should be clear from the context). We will use the normalized spherical harmonics $Y_\ell^s(\lambda, \phi)$, which play the role of sines and cosines on the sphere. The $Y_\ell^s$ are defined by

$$Y_\ell^s(\lambda, \phi) = \begin{cases} \theta_{\ell s}\cos s\lambda P_\ell^s(\sin \phi) & 0 \leq s \leq \ell \\ \theta_{\ell s}\sin s\lambda P_\ell^{|s|}(\sin \phi) & -\ell \leq s < 0, \end{cases} \quad \ell = 1, 2, \ldots$$

$$\theta_{\ell s} = \begin{cases} \sqrt{2}\,\sqrt{\dfrac{2\ell + 1}{4\pi}\dfrac{(\ell - |s|)!}{(\ell + |s|)!}} & s \neq 0 \\[2mm] \dfrac{2\ell + 1}{4\pi} & s = 0 \end{cases}$$

$$Y_0^0(\lambda, \phi) = 1$$

where the $P_\ell^s$ are the Legendre functions [33].

The spherical harmonics are the eigenfunctions of the (spherical) Laplacian

(3.1) $$\Delta Y_\ell^s = -\ell(\ell + 1)Y_\ell^s,$$

where

$$\Delta f = \frac{1}{a}\left[\frac{1}{\cos^2\phi}f_{\lambda\lambda} + \frac{1}{\cos \phi}(\cos \phi f_\phi)_\phi\right]$$

and $a$ is the radius of the sphere. The spherical harmonics form a complete orthonormal system on $L_2(S)$. If $f \in L_2(S)$, it has the Fourier-Bessel expansion

$$f \sim \sum_{\ell=0}^{\infty} \sum_{s=-\ell}^{\ell} f_{\ell s} Y_{\ell}^{s}$$

where $f_{\ell s} = \int_S f(P) Y_{\ell}^{s}(P) dP$, see [17]. Let

$$J(f) = \sum_{\ell=1}^{\infty} \sum_{s=-\ell}^{\ell} \frac{f_{\ell s}^2}{\lambda_{\ell s}}$$

where the $\{\lambda_{\ell s}\}$ are some nonnegative numbers with $|\lambda_{\ell s}| \to 0$ as $\ell \to \infty$. The set of all functions in $L_2(S)$ with $J(f)$ finite can be taken as a Hilbert space $X$ with $J^{1/2}$ as a seminorm and with the constant functions as its null space. Space $X$ can be made into a reproducing kernel Hilbert space, provided

$$\sum_{\ell=1}^{\infty} (2\ell + 1) \max_{|s| \leq \ell} \lambda_{\ell s} < \infty.$$

This follows since

$$|f(P)| \leq |f_{00}| + \sum_{\ell=1}^{\infty} \sum_{s=-\ell}^{\ell} |f_{\ell s}||Y_{\ell}^{s}(P)|$$

$$\leq |f_{00}| + \left( \sum_{\ell=1}^{\infty} \sum_{s=-\ell}^{\ell} \frac{f_{\ell s}^2}{\lambda_{\ell s}} \right)^{1/2} \left( \sum_{\ell=1}^{\infty} \sum_{s=-\ell}^{\ell} \lambda_{\ell s}[Y_{\ell}^{s}(P)]^2 \right)^{1/2}$$

$$\leq |f_{00}| + J^{1/2}(f) \frac{1}{4\pi} \sum_{\ell=1}^{\infty} (2\ell + 1) \max_{|s| \leq \ell} \lambda_{\ell s},$$

where, in obtaining the last inequality, we have applied the addition formula for spherical harmonics [17].

$$(3.2) \qquad \sum_{s=-\ell}^{\ell} Y_{\ell}^{s}(P) Y_{\ell}^{s}(P') = \frac{2\ell + 1}{2\pi} P_{\ell}(\cos \gamma(P, P')),$$

where $\gamma(P, P')$ is the angle between $P$ and $P'$, the $P_{\ell}$ are the Legendre polynomials, and $P_{\ell}(0) = 1$.

It is easy to show that if $\lambda_{\ell s} = [\ell(\ell + 1)]^{-m}$, then

$$(3.3) \quad J(f) = \begin{cases} \int (\Delta^{m/2} f)^2 dP & m \text{ even} \\ \int \{(\Delta^{(m-1)/2} f)_{\lambda}^2/\sin^2\phi + (\Delta^{(m-1)/2} f)_{\phi}^2\} \, dP & m \text{ odd}. \end{cases}$$

For $X$ to be a reproducing kernel space, it is necessary that $m > 1$ (although not necessarily an integer). Later on, we will use the fact that $(\Delta f)(P)$ will be a bounded linear functional in $X$ for each $P \in S$ provided

$$\sum_{\ell=1}^{\infty} \ell^2(\ell + 1)^2(2\ell + 1) \max_{|s| \leq \ell} \lambda_{\ell s} < \infty.$$

We remark that, if $\lambda_{/s} = \lambda_{/}$ independent of $s$, then it can be shown by the use of (3.2) that $J(\cdot)$ is isotropic; that is, it is invariant under the group of all rotations of the sphere. The seminorms $J_m^d$ of §2 are also isotropic; they are invariant under the group of all rotations and translations of $E^d$. The vector splines on the sphere described below also result from isotropic seminorms. There are, of course, situations when a specifically anisotropic seminorm may be called for (for example, when $x_1, \ldots, x_{d-1}$ are space variables and $x_d$ is a time variable, or, when the earth's Coriolis force is an important factor) but we omit detailed discussion of that case here.

We are now ready to define vector splines on the sphere. Let $\mathbf{V} = (U, V)$ be a sufficiently regular horizontal vector field on the sphere, where $U(P)$ is the eastward component and $V(P)$ is the northward component at $P$. The vorticity $\zeta$ and the diverence $D$ of $\mathbf{V}$ are given by

$$(3.4a) \qquad \zeta = \frac{1}{a \cos \phi} \left[ -\frac{\partial}{\partial \phi} (U \cos \phi) + \frac{\partial V}{\partial \lambda} \right]$$

and

$$(3.4b) \qquad D = \frac{1}{a \cos \phi} \left[ \frac{\partial U}{\partial \lambda} + \frac{\partial}{\partial \phi} (V \cos \phi) \right].$$

Then there exists (by Helmoltz' Theorem) two functions $\Psi(P)$ and $\Phi(P)$, $P \in S$, called the stream function and the velocity potential respectively, with the following properties:

$$(3.5) \quad U = \frac{1}{a} \left( -\frac{\partial \Psi}{\partial \phi} + \frac{1}{\cos \phi} \frac{\partial \Phi}{\partial \lambda} \right), \qquad V = \frac{1}{a} \left( \frac{1}{\cos \phi} \frac{\partial \Psi}{\partial \lambda} + \frac{\partial \Phi}{\partial \phi} \right)$$

and

$$(3.6) \qquad \zeta = \Delta \Psi, \qquad D = \Delta \Phi.$$

Functions $\Psi$ and $\Phi$ are uniquely determined up to a constant, which we will take to be determined by $\int_S \Psi(P) dP = \int_S \Phi(P) dP = 0$.

Let $X$ be the collection of all pairs $(\Psi, \Phi)$ on the sphere which integrate to zero, are square integrable and

$$J^{(1)}(\Psi) = \sum_{/=1}^{\infty} \sum_{s=-/}^{/} \frac{\Psi_{/s}^2}{\lambda_{/s}(1)} < \infty, \qquad \Psi_{/s} = \int \Psi(P) Y_/^s(P) dP$$

$$J^{(2)}(\Phi) = \sum_{/=1}^{\infty} \sum_{s=-/}^{/} \frac{\Phi_{/s}^2}{\lambda_{/s}(2)} < \infty, \qquad \Phi_{/s} = \int \Phi(P) Y_/^s(P) dP$$

where $\{\lambda_{/s}(1)\}$ and $\{\lambda_{/s}(2)\}$ are strictly positive sequences satisfying

$$\sum_{/=1}^{\infty} /^2(/ + 1)^2 (2/ + 1) \max_s \lambda_{/s}(i) < \infty, \quad i = 1, 2.$$

Space $X$ is clearly a Hilbert space with square norm $\|(\Psi, \Phi)\|^2 = J^{(1)}(\Psi) + (1/\delta)J^{(2)}(\Phi)$ for any fixed $\delta > 0$ and furthermore both members of each pair possess Laplacians everywhere.

Let data $(U_i, V_i)$ be given, where $U_i$ and $V_i$ are supposed to be noisy measurements on the true field $(U(P_i), V(P_i))$. Let $(\Psi, \Phi)$ be the stream function and velocity potential associated with $(U, V)$ and let $(\Psi_{\lambda,\delta}, \Phi_{\lambda,\delta})$ be the minimizer of

$$
(3.7) \quad \frac{1}{n}\sum_{i=1}^{n}\left(-\frac{1}{a}\frac{\partial\Psi}{\partial\phi}(P_i) + \frac{1}{a\cos\phi_i}\frac{\partial\Phi}{\partial\lambda}(P_i) - U_i\right)^2
$$

$$
+ \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{a\cos\phi_i}\frac{\partial\Psi}{\partial\lambda}(P_i) + \frac{1}{a}\frac{\partial\Phi}{\partial\phi}(P_i) - V_i\right)^2 + \lambda[J_1(\Psi) + \frac{1}{\delta}J_2(\Phi)].
$$

Note that in the residual sum of squares above, $U(P_i)$ and $V(P_i)$ are expressed in terms of $\Psi$ and $\Phi$ via (3.5). A unique minimizer $(\Psi_{\lambda,\delta}, \Phi_{\lambda,\delta})$ exists for each $\lambda > 0$, $\delta > 0$ and the resulting wind field $(U_{\lambda,\delta}, V_{\lambda,\delta})$ constructed from $(\Psi_{\lambda,\delta}, \Phi_{\lambda,\delta})$ may be termed a vector smoothing spline field. Its vorticity and divergence will be given by $\zeta_{\lambda,\delta} = \Delta\Psi_{\lambda,\delta}$, $D_{\lambda,\delta} = \Delta\Phi_{\lambda,\delta}$.

The parameter $\delta$ influences the relative amount of energy that will be assigned to the divergent and nondivergent part of the estimated vector field. It is tempting to set $\delta = 1$, however, in some applications this may be a priori not the correct value. (The application we have in mind is to upper altitude horizontal wind fields in mid latitudes, where the divergent part is generally smaller than the nondivergent part.) In the numerical study described below we have investigated the ability of the GCV method to estimate a good value of $\delta$ as well as a good value of $\lambda$.

Explicit expressions for the minimizer of (1.1) with $J$ given by (3.3) and related expressions may be obtained in terms of infinite series, and various methods of approximating the solution by a finite form may be based on approximations to the relevant r.k. See [27, 8, 9, 18]. Wendelberger [32] has found a closed form expression for the cases of (3.3) with $m = 2$ and $3$ in terms of $di$- and tri-logarithms. These results could have been extended to aid in the minimization of (3.7) for special choices of the $\lambda_{/s}(i)$. However, in this work we have chosen to obtain the minimizer of (3.7) in the span of two sets of the $\tilde{N} = N(N + 2)$ spherical harmonics. This proved to be quite feasible for $N$ up to around 16 on the Amdahl at Goddard Space Flight Center.

Let

$$
(3.8) \quad \Psi = \sum_{/=1}^{N}\sum_{s=-/}^{/}\alpha_{/s}Y_/^s
$$

$$
\Phi = \sum_{/=1}^{N}\sum_{s=-/}^{/}\beta_{/s}Y_/^s.
$$

The coefficients $\{\alpha_{\prime s}\}$ and $\{\beta_{\prime s}\}$ for which (3.7) is minimized, and the cross-validation estimates of $\lambda$ and $\delta$ are found as follows.

First, the indices $(\ell, s)$, $s = -\ell, \ldots, \ell$, are renumbered from 1 to $\tilde{N}$. Let $X_\phi$ be the $n \times \tilde{N}$ matrix with $(i, \ell s)$-th entry

$$\frac{1}{a} \frac{\partial}{\partial \phi} Y_\ell^s(P_i)$$

and $X_\lambda$ be the $n \times \tilde{N}$ matrix with $(i, \ell s)$-th entry

$$\frac{1}{a} \frac{1}{\cos \phi_i} \frac{\partial}{\partial \lambda} Y_\ell^s(P_i)$$

and let $X$ be the $2n \times 2\tilde{N}$ matrix

$$(3.9) \qquad X = \begin{bmatrix} -X_\phi & X_\lambda \\ X_\lambda & X_\phi \end{bmatrix}.$$

Let $D_\delta$ be the $2\tilde{N} \times 2\tilde{N}$ matrix

$$(3.10) \qquad D_\delta = \begin{bmatrix} D_1 & 0 \\ 0 & \delta D_2 \end{bmatrix}$$

where $D_i$ is the $\tilde{N} \times \tilde{N}$ diagonal matrix with $(\ell s, \ell s)$-th entry $\lambda_{\ell s}(i)$, $i = 1, 2$. Letting $z = (U_1, \ldots, U_n, V_1, \ldots, V_n)$, $\gamma = (\alpha_1, \ldots, \alpha_{\tilde{N}}, \beta_1, \ldots, \beta_{\tilde{N}})$, it is seen by substituting (3.8) into (3.7) that we have to find $\gamma$ which minimizes

$$\frac{1}{n} \|z - X\gamma\|^2 + \lambda\gamma' D_\delta^{-1}\gamma.$$

The minimizer is

$$(3.11) \qquad \gamma = (X'X + n\lambda D_\delta^{-1})^{-1}X'z.$$

By the use of (3.5) and (3.11), it follows that the estimated wind field $(U_{\lambda,\delta}, V_{\lambda,\delta})$ at the data points satisfies

$$(3.12) \qquad \begin{bmatrix} U_{\lambda,\delta}(P_1) \\ \vdots \\ Y_{\lambda,\delta}(P_n) \\ V_{\lambda,\delta}(P_1) \\ \vdots \\ V_{\lambda,\delta}(P_n) \end{bmatrix} = A(\lambda, \delta)z$$

where $A(\lambda, \delta)$ is the $2n \times 2n$ "influence" matrix $A(\lambda, \delta) = X(X'X + n\lambda D_\delta^{-1})^{-1}X'$. The cross-validation function $V(\lambda, \delta)$ to be minimized in $\lambda$ and $\delta$ is

$$(3.13) \qquad V(\lambda, \delta) = \dfrac{\dfrac{1}{n} \|(I - A(\lambda, \delta))z\|^2}{\dfrac{1}{n} [\mathrm{Tr}(I - A(\lambda, \delta))]^2}.$$

The minimizer of $V(\lambda, \delta)$ was found, in the study described below, as follows. For fixed $\delta$, let $W_\delta = XD_\delta^{1/2}$, and let the singular value decomposition ($SVD$) of $W_\delta$ be

$$(3.14) \qquad W_\delta = UD_W V'$$

where $UU' = U'U = I_{2n \times 2n} = V'V$ and $D_W$ is a diagonal matrix with entries $b_1, \ldots, b_{2n}$. (Here, $n \leq \tilde{N}$.) $U$, $\{b_i\}$ and $V'$ are computed using LINPACK. Letting $w = (w_1, \ldots, w_{2n})' = U'z$, then

$$(3.15) \qquad V(\lambda, \delta) = \dfrac{\dfrac{1}{2n} \displaystyle\sum_{i=1}^{2n} \left( \dfrac{n\lambda}{b_\nu^2 + n\lambda} \right)^2 w_\nu^2}{\left( \dfrac{1}{2n} \displaystyle\sum_{\nu=1}^{2n} \dfrac{n\lambda}{b_\nu^2 + n\lambda} \right)^2}$$

and

$$(3.16) \qquad \gamma = D_\delta^{1/2} V \begin{bmatrix} \dfrac{b_1}{b_1^2 + n\lambda} & & 0 \\ & \ddots & \\ 0 & & \dfrac{b_{2n}}{b_{2n}^2 + n\lambda} \end{bmatrix} w.$$

For fixed $\delta$, $\hat{\lambda}(\delta)$, the minimizer of (3.15) was easily found by a global search in increments of $\log \lambda$. Then $V(\hat{\lambda}(\delta), \delta)$ was plotted for eight values of $\delta$ chosen in powers of one-sixth, and the minimum was readily evident. No doubt more efficient and automatic search procedures can be found. However, $V$ is not a convex function of $\lambda$, and it is possible to encounter more than one local minimum. We have on occasion seen this when analyzing experimental data with small $n$.

For large $n$, $N$, and $W_\delta$ poorly conditioned, computing the SVD can be expensive, or it can fail to converge in a reasonable time. Some shortcut methods which alleviate this problem somewhat and use less storage have been developed. See [1].

We designed a Monte Carlo experiment to estimate the accuracy with which the method is able to estimate vorticity and horizontal divergence of the 500 mb. horizontal wind field over North America from observations on the East ($U_i$) and North ($V_i$) components of the wind taken by the North American radiosonde network. We will show the results of one such experiment. This information is useful in meteorological studies. Since the accuracy will depend on the distribution of data points, and the relations between the energy in the true divergent and nondivergent parts

of the wind field as well as the measurement error, pains were taken to make the simulated truth and simulated measurment errors as realistic as possible. Nevertheless, the results obtained below must be treated as a lower bound to accuracy expected in practice. In particular, it will be seen below that a particular sequence of $\{\lambda_{\ell s}(i)\}$ has been chosen to define $J^{(1)}$ and $J^{(2)}$, based on some (partly ad hoc) meteorological considerations, and the simulated "truth" below was generated via a random model based on the same $\{\lambda_{\ell s}(i)\}$. In practice the matching of the $\{\lambda_{\ell s}(i)\}$ used in $J_1$ and $J_2$ cannot be expected to match nature as well. (Also energy exists in nature at higher wavenumbers than are being simulated.)

We digress briefly to discuss the choice of the $\lambda_{\ell s}(i)$, $i = 1, 2$, in this study. Suppose a (univariate) function on the sphere can be considered to be a random linear combination of spherical harmonics $f(P) = \sum\sum f_{\ell s}Y_{\ell}^s(P)$ where the $f_{\ell s}$ are supposed to be realizations of zero mean independent (Gaussian) random variables with $Ef_{\ell s}^2 = b\lambda_{\ell s}$ for some $b$.

Letting $y_i = f(P_i) + \varepsilon_i$ where the $\{\varepsilon_i\}$ are independent and normally distributed with mean zero and variance $\sigma^2$, then $f_\lambda$ is the Bayes estimate of $f$; that is $f_\lambda(P) = E\{f(P) | y_1, \ldots, y_n\}$ where $f_\lambda$ is the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}(u_i - f(P_i))^2 + \lambda J(f),$$

where $\lambda = \sigma^2/nb$ and $J(f) = f_{\ell s}^2/\lambda_{\ell s}$. For further details, see [15, 24, 30].

With many (possibly oversimplifying) assumptions information concerning $\{\lambda_{\ell s}\}$ may be obtained from historical records. For example, suppose $\lambda_{\ell s} = \lambda_{\ell}$, independent of $s$ (that is, the resulting $J$ will be isotropic). Then

$$Ef(P)f(P') = E\sum_{\ell s}\sum_{\ell' s'} f_{\ell s}f_{\ell' s'}\, Y_{\ell}^s(P)Y_{\ell'}^{s'}(P')$$

$$= b\sum_{\ell s}\lambda_{\ell}Y_{\ell}^s(P)Y_{\ell}^s(P')$$

$$= b\sum_{\ell}\lambda_{\ell}\frac{2\ell+1}{4\pi}P_{\ell}(\cos\gamma(P, P'))$$

$$\doteq r(\gamma(P, P')).$$

Figure 3.1 gives a sequence of idealized $\{\lambda_{\ell}\}$ of which the first fourteen were used in this study and Figure 3.2 gives a plot of the correlation function $\rho(\gamma) = r(\gamma)/r(0)$ associated with the $\{\lambda_{\ell}\}$ of Figure 3.1. Figure 3.3 gives a fitted isotropic correlation function estimated from historical 500 mb. geopotential fields by Julian and Thiebaux [12] which roughly matches that of Figure 3.2. Although in meteorological practice the isotropy and other assumptions made below may be suspect, in principle appropriate sequences $\{\lambda_{\ell s}\}$, or at least appropriate rates of decay of the $\{\lambda_{\ell s}\}$ may be obtained from historical records. See, for example, [13].
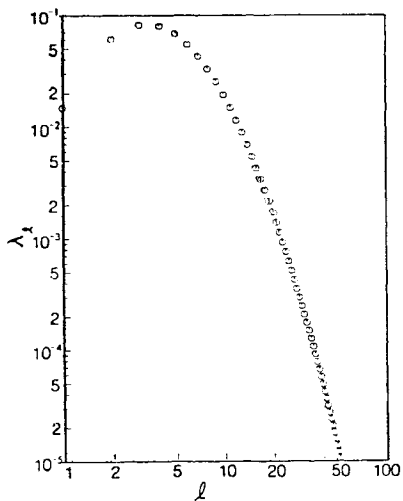
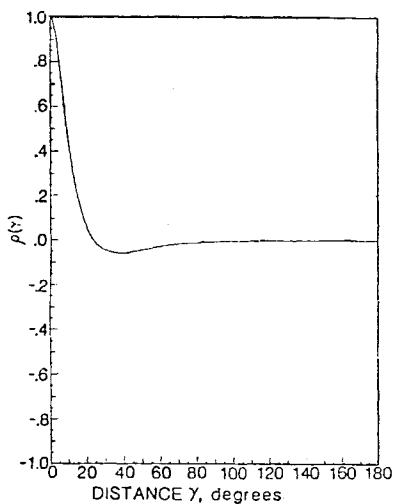FIGURE 3.1. Idealized $\lambda_\ell$.

FIGURE 3.2. Correlation function for
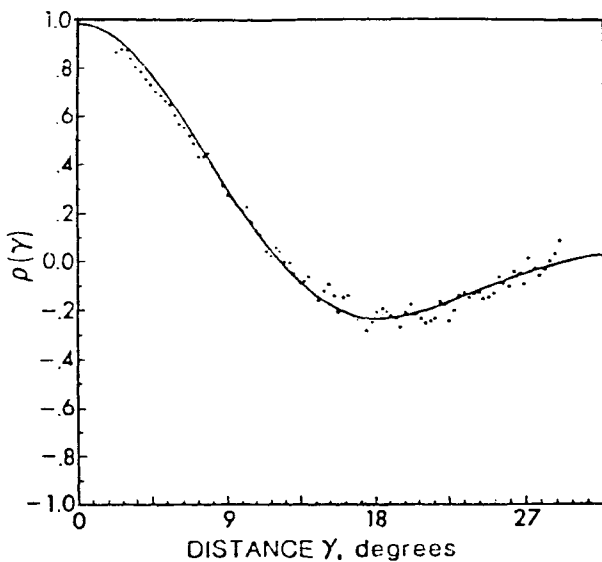the $(\lambda_\ell)$ of Figure 3.1.
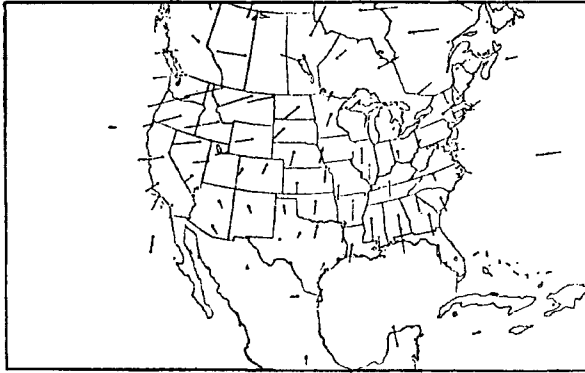


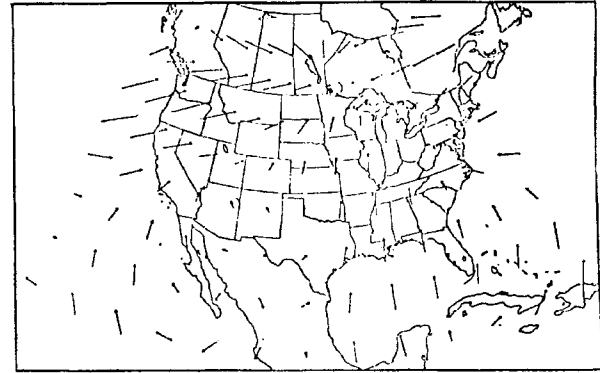FIGURE 3.3. Sample Correlation Function.

FIGURE 3.4.  Simulated Wind Data.



FIGURE 3.5.  Estimated Wind Field.
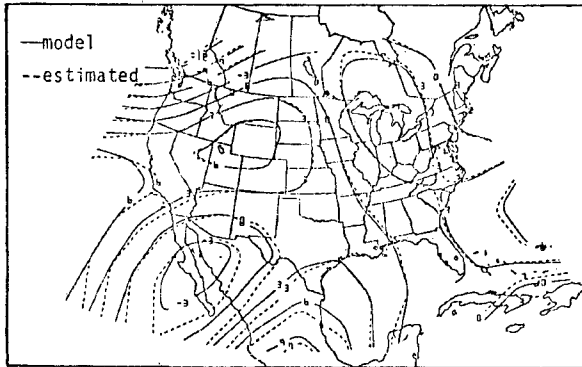


FIGURE 3.6   Model and Extimated
Vorticity, $X$ $10^{-5}$/sec.
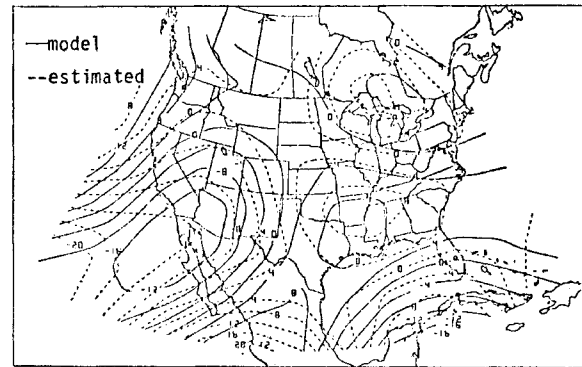


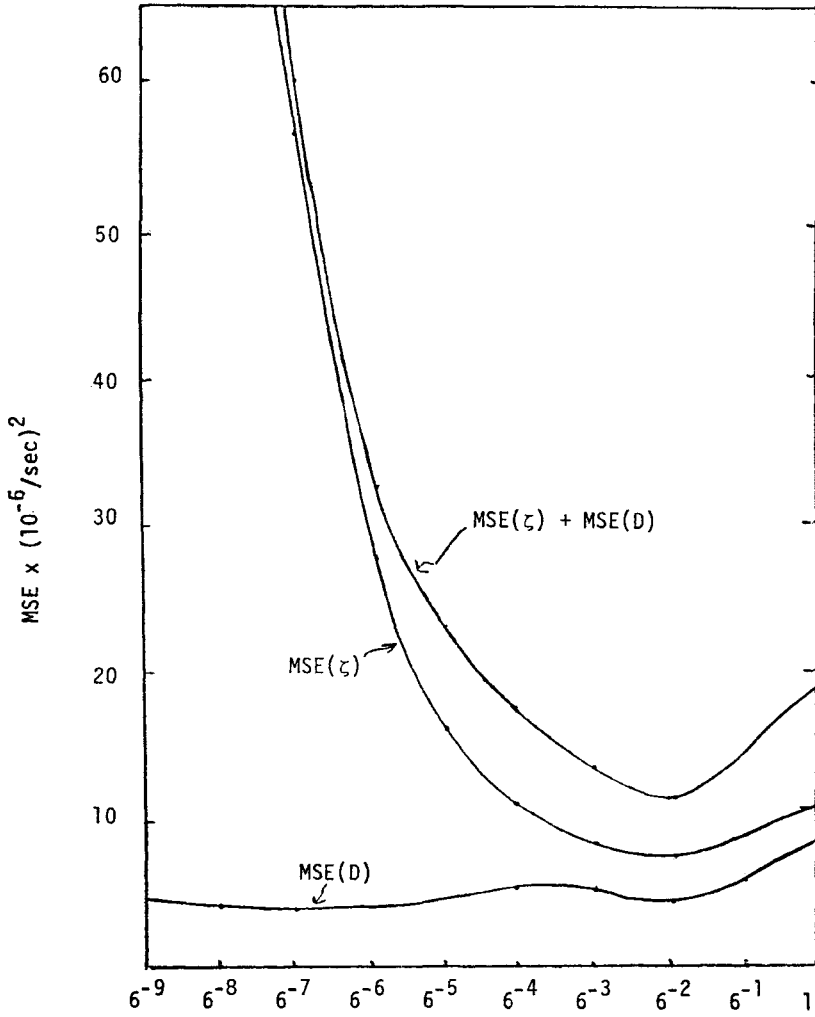FIGURE 3.7.  Model and Estimated
Divergence, $X$ $10^{-6}$/sec.

FIGURE 3.8 Mean square error in the estimated voriticty and divergence.

We obtained a model stream function and velocity potential of the form

(3.17)

$$\Psi = C_1 \sum_{l=1}^{N} \sum_{s=-l}^{l} a_{ls} Y_l^s$$

$$\Phi = C_2 \sum_{l=1}^{N} \sum_{s=-l}^{l} b_{ls} Y_l^s$$

by choosing $a_{ls}$ and $b_{ls}$ as normally distributed pseudo-random numbers with mean zero and variances $\lambda_{ls}(1) = \lambda_{ls}(2) = \lambda_l$ given in Figure 3.1. Parameters $C_1$ and $C_2$ were scale factors chosen so that the simulated $\zeta = \Delta\Psi$ and $D = \Delta\Phi$ had magnitudes typical of real 500 mb. wind fields.

$$\frac{1}{4\pi}\left(\int \zeta^2 \, dP\right)^{1/2} = 6 \times 10^{-5}/\text{sec.} \qquad \left(\frac{1}{4\pi}\int D^2 dP\right)^{1/2} = 1 \times 10^{-5}/\text{sec.}$$

If $C_1 = C_2$, then the optimal $\delta$ would be one (or near to one). However, here, the divergent part of the wind field is smaller than the nondivergent part, and the optimal $\delta$ will be less than one. The experiment below tests the ability of GCV to estimate a good value of $\delta$ (as well as $\lambda$).

Model wind vectors $(U(P_i), V(P_i))$ were computed from the model $(\Psi, \Phi)$ of (3.17) for $\{P_i\}$ corresponding to $n = 114$ North American radiosonde stations. The data $z = (U_1, \ldots, U_n, V_1, \ldots, V_n)$, where $U_i = U(P_i) + \varepsilon_i^U$, $V_i = V(P_i) + \varepsilon_i^V$, were constructed by adding the measurement errors $\varepsilon_i^U$, $\varepsilon_i^V$ as normally distributed pseudo-random numbers with mean zero and standard deviation $\sigma = 2.5$ meters/sec., a realistic value for the measurement error standard deviation.

Figure 3.4 shows the simulated wind vectors. Figure 3.5 shows the estimate of the true wind field, plotted on a $5° \times 5°$ grid in latitude and longitude. Figures 3.6 and 3.7 show the model and estimated vorticity and divergence, respectively. In Figures 3.5–3.7 $\delta = 1/36$ was used, which was the minimizer of $V(\hat{\lambda}(\delta), \delta)$ found by the search described above. Figure 3.8 gives $\text{MSE}(\zeta_{\hat{\lambda}(\delta),\delta})$ and $\text{MSE}(D_{\hat{\lambda}(\delta),\delta})$ and their sum, where

$$\text{MSE}(\zeta_{\lambda,\delta}) = \frac{1}{K} \sum_{k=1}^{K} (\zeta_{\lambda,\delta}(P_k) - \zeta(P_k))^2$$

$$\text{MSE}(D_{\lambda,\delta}) = \frac{1}{K} \sum_{k=1}^{K} (D_{\lambda,\delta}(P_k) - D(P_k))^2.$$

The $\{P_k\}$ constituted a regular grid inside the United States. It can be seen from Figure 3.8 that if $\delta$ is taken as too small (i.e., divergence is suppressed), then the mean square error in the estimated vorticity becomes large, and similarly if $\delta$ is too large, then the mean square error in the estimated divergence becomes large. It appears that the GCV estimate of $\delta$ here is quite close to the $\delta$ which minimizes $\text{MSE}(\zeta) + \text{MSE}(D)$.

## References

1. D. Bates and G. Wahba, *Computational methods for generalized cross-validation with large data sets*, in *Treatment of Integral Equations by Numerical Methods*, C. T. H. Baker and G. F. Miller, eds., Academic Press, 1982, 283–296.

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

3. C. deBoor, *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.

4. J. J. Dongarra, C. B. Moler, J. R. Bunch and G. W. Stewart, LINPACK Users' Guide, SIAM, Philadelphia, 1979.

5. J. Duchon, *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces*, RAIRO Anal. Number. 10 (1976), 5–12.

6. ———, *Splines minimizing rotation—invariant semi-norms in Sobolev spaces*, in Constructive Theory of Functions of Several Variables, W. Schempp and K. Zeller, eds., Springer-Verlag, Berlin, 1977.

7. N. Dyn and D. Levin, *A procedure for generating diagonal dominance in ill-conditioned systems originating from integral equations and surface interpolation*, School of Mathematical Sciences, Tel Aviv University, TR #81–18, 1981.

8. W. Freeden, *On spherical spline interpolation and approximation*, Math. Meth. in The Appl. Sci. 3 (1981), 551–575.

9. ———, *On approximation by harmonic splines*, Manuscripta Geodaetica 6 (1981), 193–244.

10. H. A. Gamber, *Choice of an optimal shape parameter when smoothing noisy data*, Commun. Statist, A8 (1979), 14.

11. G. Golub, M. Heath and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics 21 (1979), 214–223.

12. P. R. Julian and M. J. Thiebaux, *On some properties of correlation functions used in optimum interpolation schemes*, Monthly Weather Review, 103 (1975), 605–616.

13. A. Kasahara, *Normal modes of ultra-long waves in the atmosphere*, Monthly Weather Review 104 (1976), 669–690.

14. G. Kimeldorf and G. Wahba, *A correspondence between Bayesian estimation of stochastic processes and smoothing by splines*, Ann Math. Statist. 41 (1970).

15. ———, *Some results on Techbycheffian spline functions*, J. Math. Anal. and Applic. 33 (1971).

16. J. Meinguet, *Multivariate interpolation at arbitrary points made simple*, J. Appl. Math. Phys. 30 (1979).

17. G. Sansone, *Orthogonal Functions*, revised English edition, Interscience, N.Y., 1959.

18. L. Shure, R. L. Paker and G. E. Backus, *Harmonic splines for geomagnetic modelling*, Physics of the Earth and Planetary Interiors 28 (1982).

19. P. Speckman, *Efficient nonparametric regression with cross-validated smoothing splines*, manuscript, 1982.

20. F. Uteras, *Quelques résultats d'optimalité pour la methode de validation croisée*, Seminiaire d'Analyse Numerique, Mathématiques Appliquées, Université Scientifique et Médicale de Grenoble, No. 301, 1978.

21. ———, *Cross-validation techniques for smoothing spline functions in one or two dimensions*, in Smoothing Techniques for Curve Estimation, T. Gasser and M. Rosenblatt, eds., Lecture Notes in Math., 757, Springer-Verlag, Berlin, 1979.

22. ——, *Optimal smoothing of noisy data using spline functions*, SIAM J. Sci. Stat. Comput. **2** (1981), 349–362.

23. G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numerical Analysis, **14** (1977).

24. ——, *Improper priors, spline smoothing and the problem of guarding against model errors in regression*, J. Royal Stat. Soc., Ser. B., **40** (1978).

25. ——, *Spline bases, regularization, and generalized cross-validation for solving approximation problems with large quantities of noisy data*, Proceedings of the International Conference on Approximation Theory in Honor of George Lorenz, Ward Cheney, ed., Academic Press, NY, 1980.

26. ——, *Ill-posed problems: Numerical and statistical methods for midly, moderately, and severely ill-posed problems with noisy data*, University of Wisconsin-Madison Statistics Department Technical Report No. 595, February 1980, to appear in the Proceedings of the International Conference on Ill-Posed Problems, M.Z. Nashed, ed.

27. ——, *Spline interpolation and smoothing on the sphere*, SIAM J. Scientific and Statical Computing **2** (1981).

28. ——, *Data-based optimal smoothing of orthogonal series density estimates*, Ann. Statist. **9** (1981).

29. ——, *Constrained regularization for ill-posed linear operator equations with applications in meteorology and medicine*, in Statistical Decision Theory and Related Topics III, Vol. 2, S. S. Gupta and J. O. Berger, eds., Academic Press, NY, 1982, 383–418.

30. ——, *Vector splines on the sphere with application to the estimation of vorticity and divergence from discrete, noisy data*, University of Wisconsin-Madison Statistics Department Technical Report No. 674, May 1982, Multivariate Approximation Theory, Vol. 2, W. Schempp and K. Zeller, eds., Birkhauser Verlag.

31. —— and J. Wendelberger, *Some new mathematical methods for variational objective analysis using splines and cross-validation*, Monthly Weather Review **108** (1980), 36–57.

32. J. Wendelberger, *Smoothing noisy data with multidimensional splines and generalized cross-validation*, Ph.D. Thesis, Department of Statistics, University of Wisconsin-Madison, 1982.

33. E. T. Whittaker and G. N. Watson, *A Course on Modern Analysis*, Cambridge University Press, Cambridge, 1958.

DEPARTMENT OF STATISTICS, UNIVERSITY OF WISCONSIN, MADISON, WI 53706