
DEPARTMENT OF STATISTICS

University of Wisconsin
1210 West Dayton Street
Madison, WI 53706

TECHNICAL REPORT NO. 755

December 1984

VARIATIONAL METHODS FOR MULTIDIMENSIONAL
INVERSE PROBLEMS

by

Grace Wahba

Prepared for the Proceedings of the Workshop on Advances in Remote Sensing
Retrieval Methods, held at Williamsburg, VA, 30 October - 2 November, 1984,
A. Deepak, editor.

VARIATIONAL METHODS FOR MULTIDIMENSIONAL INVERSE PROBLEMS

Grace Wahba
Department of Statistics
University of Wisconsin-Madison
1210 W. Dayton St.
Madison, WI 53706

ABSTRACT

By combining ideas from various sources, we propose a class of physical variational methods for estimating three dimensional temperature structure from satellite radiance data using many sets of vertical and/or oblique soundings simultaneously. The class of methods proposed takes advantage of assumed slowly varying temperature structure in the horizontal and linearizes the inherently (mildly) nonlinear problem of temperature retrievals as late as possible. The method can handle irregularly spaced clear column radiances. Methods are proposed for the specific inclusion and weighting of forecast information, radiosonde data and tropopause height information. Following Smith (1984) it is suggested how temperature and water vapor may be simultaneously analyzed with the proposed methods. So long as nature is "smooth" or highly correlated it is in principle advantageous to build this "smoothness" or correlation information into the analysis, via analyzing large quantities of data simultaneously. A method for using GCV (generalized cross validation) to get (some of) the tuning parameters adaptively in nonlinear and constrained problems with large data sets is described. The main drawback to dealing with large sets of data simultaneously is the computational cost in time and storage. Various algorithms and shortcuts are proposed for approaching the computational problem.

1. INTRODUCTION

The approach to non-linear physical variational analysis of satellite radiance data that is the subject of this paper may be found in O'Sullivan (1983) and O'Sullivan and Wahba (1984) (O'S & W). This approach is a member of the class of methods known in the approximation theory literature as regularization methods. Related ideas have been discussed by Hoffman (1983,1984). The extension here to two and three dimensions of the one dimensional retrieval method of O'S & W may be thought of as a form of "satellite tomography" as described in Fleming (1983). The approach to combining forecast and observation is along the lines proposed in Wahba (1982c) and is not unrelated to Kalman filtering. The class of methods that we propose are applicable to other remote sensing problems, however, this paper is written with the specific application to three dimensional temperature retrievals in mind.

Supported by NASA (Grant NAG5-316) and ONR (Contract N00014-77-C-0675).

In this work we are going to ignore the very important problems of systematic or bias errors in the atmospheric transmittance functions and elsewhere. See Fleming and Crosby (1983), Fleming, Crosby and Neuendorffer (1983), Susskind, Rosenfield and Reuter (1983) and O'Sullivan (1983) for discussions of bias problems. We are also going to ignore most of the problems associated with clouds (see Susskind et al (1983,1984), Hoffman (1983)). We will assume that clear (cloud free) columns can be identified, however, and our three dimensional retrievals will in principle be able to take advantage of scattered clear column radiances even though their spacing may be fairly irregular. The method is expected to provide a reasonable interpolation between clear column data.

After a brief introduction to radiance data, we will first describe the O'S & W physical variational method for single column (one dimensional) temperature retrievals. Then we show how this method may be extended to use many scattered clear column radiances to get a three dimensional temperature analysis. Next we discuss how forecast, radiosonde and tropopause height information can be simultaneously incorporated into the analysis. We briefly suggest how water vapor and temperature may be simultaneously analyzed within the present class of methods. We conclude with a few remarks concerning numerical methods for solving the large and huge variational problems that result from the proposed methods.

2. SATELLITE RADIANCE DATA

A (single column of) satellite radiance observations may be modelled as

$$y_i = N_i(T) + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where N_i is a non-linear functional of the temperature T , and

$$N_i(T) = \int f_i(\nu) R_\nu d\nu \quad (2)$$

where $f_i(\nu)$ is the instrument response of the i th channel to the incident radiation R_ν at frequency ν , where

$$R_\nu(\theta) = \epsilon_\nu(\theta) B_\nu(T_s) \tau_\nu(p_s, \theta) + \int B_\nu[T(p, p)] \frac{\partial \tau_\nu(p, \theta)}{\partial \ln p} d \ln p \quad (3)$$

$\epsilon_\nu(\theta)$ is the emissivity of the surface s at zenith angle θ , $B_\nu(T)$ is the Planck function for emitted radiance of a blackbody at frequency ν and temperature T given by

$$B_\nu(T) = c_1 \nu^3 / \{\exp[c_2 \nu/T] - 1\} \quad (4)$$

(where c_1 and c_2 are constants), and $\tau_\nu(p, \theta)$ is the atmospheric

transmittance from pressure p to the top of the atmosphere at zenith angle θ . The integral is taken along the line of sight, which depends on the subsatellite point and the look direction. Reflected radiation is being ignored. See Fritz et al (1972), Susskind et al (1984) for further details. For the purposes of this paper we will assume that the response function f is narrow. With these simplifying assumptions the channel response is absorbed into the atmospheric transmittance function, giving

$$N_i(T) = \int B_v[T(p)]K_{v,\theta}(p)dp \quad (5)$$

for "known" $K_{v,\theta}$ where for simplicity (only) we are assuming that surface quantities have been subtracted out.

More generally if ϵ is known the surface temperature can be included as part of the analysis and will appear in the definition of N_i . In addition the transmittance has a (usually ignored) dependence in the unknown T . In a high precision method, we would write $K_{v,\theta}(T(p),p)dp$ in (5) instead of $K_{v,\theta}(p)$. The transmittance may also be allowed to depend on water vapor, see Section 7.

3. THE PHYSICAL VARIATIONAL METHOD OF O'S & W

The retrieval algorithm proposed by O'S & W finds the temperature analysis T as the minimizer (in an appropriate class of functions⁺) of

$$\frac{1}{2} \sum_{i=1}^n w_i (y_i - N_i(T))^2 + \lambda J(T) \quad (6)$$

where $J(T)$ is a smoothness penalty and the w_i are inversely proportional to the mean square values of the ϵ_i . When the smoothness penalty is a nonnegative quadratic form then the choice of smoothness penalty is equivalent to treating T as though T were a (Gaussian) stochastic process with a given prior covariance function. (We will return to this point with some examples later). O'S & W used the smoothness penalty

⁺If desired, we can expand about climatology T_0 so that $T = T_0 + \delta$

and solve for δ . In that case, the smoothness penalty will be applied to δ .

$$J(T) = \int (T''(p))^2 dp. \quad (7)$$

To avoid further notation we will assume that the w_i are absorbed into y_i and N_i , that is, $y_i/\sqrt{w_i}$ and $N_i/\sqrt{w_i}$ are the "data" and "observational functionals" respectively.

O'Sullivan (1983) discusses the exact minimization of (6) in an appropriate Hilbert space with J given by (7), but for practical purposes, it is possible to obtain essentially the same result by a minimizing (6) in a carefully chosen (sufficiently large) finite dimensional subspace spanned by appropriately chosen basis functions,

B_1, \dots, B_r . Thus we seek $T(p)$ of the form⁺

$$T(p) = \sum_{j=1}^r \beta_j B_j(p) \quad (8)$$

to minimize (6). Substituting (8) into (6) and writing

$$N_i(\beta) = N_i\left(\sum_{j=1}^r \beta_j B_j(p)\right) \text{ gives the problem}$$

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^n (y_i - N_i(\beta))^2 + \lambda \beta' \Sigma \beta \quad (9)$$

where the jk th entry σ_{jk} of Σ is given by

$$\sigma_{jk} = \int B_j''(p) B_k''(p) dp. \quad (10)$$

With the penalty functional of (7), it is natural to let the B_j be a set of cubic B-splines as discussed e.g., in deBoor (1978), see also Section 4 below.

The problem now is to choose λ and find the β which minimizes (9). The bandwidth parameter λ governs the tradeoff between the smoothness of the solution and the fidelity to the data. The choice of the parameter λ is well known to be "important" and the generalized cross validation (GCV) method is known to be an effective method for choosing λ and other certain other tuning parameters from the data. Hoffman (1984) argues in the variational problems he tried that a factor of 2 in λ is not important in terms of changing the solution but a change in the order of magnitude

⁺(T_0 may be added to the right hand side of (8), to avoid cumbersome notation we will omit this in the remainder of this paper.)

of λ will change the result. This is not at odds with our own experience. To be precise, when there is a single tuning parameter λ as we have here, a change in λ is large or small in a practical sense if it is large or small in relation to the natural eigenvalues of the problem. See Wahba (1983a) especially Fig. 2 for an explanation of this point. Hoffman tuned his variational problem by trial and error. In this paper we will assume that it is desirable to use GCV to retune the method using the data (at least occasionally) and aim towards providing cheaper approximate methods for implementing it, on large data sets, and on nonlinear and or constrained problems. Thus the algorithms proposed here are to some extent motivated by the requirement that they are compatible with the use of GCV. We believe that in any situation where repetitive analysis of data sets from changing atmospheric conditions is to be analyzed, it is appropriate to check the costs of automatic tuning against the possible benefits.

Properties of GCV are fairly well known in the case that N_i is linear in β , see Craven and Wahba (1979), Wahba and Wendelberger (1980), Wahba (1977). To motivate the generalization of GCV to the nonlinear case we review its operation in the linear case. In the linear case we have

$$N_i(\beta) = \sum_{j=1}^r x_{ij} \beta_j \quad (11)$$

for some x_{ij} . Letting X be the $n \times r$ matrix with ij th entry x_{ij} , then (9) becomes

$$\frac{1}{n} ||y - X\beta||^2 + \lambda \beta' \Sigma \beta \quad (12)$$

where $||\cdot||$ is the Euclidean norm. The minimizer β_λ of (12) is given by

$$\beta_\lambda = (X'X + n\lambda \Sigma)^{-1} X'y. \quad (13)$$

The influence matrix $A(\lambda)$ which relates the data vector y to the predicted data vector $\hat{y} = X\beta_\lambda (=A(\lambda)y)$ is

$$A(\lambda) = X(X'X + n\lambda \Sigma)^{-1} X' \quad (14)$$

and the GCV estimate of λ is given by the minimizer of $V(\lambda)$ given by

$$V(\lambda) = \frac{\frac{1}{n} \text{RSS}(\lambda)}{\left[\frac{1}{n} \text{Trace}(I-A(\lambda)) \right]^2} \quad (15)$$

where $\text{RSS}(\lambda)$ is the residual sum of squares when λ is used, and is given by

$$\text{RSS}(\lambda) = ||(I-A(\lambda))y||^2. \quad (16)$$

Background on the extension of GCV to nonlinear problems such as the one described here may be found in Wahba (1981a) and O'Sullivan (1983). The extension implemented in O'S & W which we will describe here is obtained by observing that the i th entry $a_{ii}(\lambda)$ of $A(\lambda)$ is given by

$$\frac{\partial \hat{y}_i}{\partial y_i} = a_{ii}(\lambda) \quad (17)$$

and the right hand side is independent of y_i since the relationship between \hat{y}_i and y is linear. In the nonlinear case O'S & W let $a_{ii}(\lambda)$

be an approximation to $\frac{\partial \hat{y}_i}{\partial y_i}$ evaluated at the minimizer β_λ of

of (9). This is obtained as follows: First, for fixed λ , the minimizer β_λ is obtained (approximately!) by an iterative method (to be described). Then, take the Taylor series expansion of $N_i(\beta)$ about β_λ , viz:

$$N_i(\beta) \approx N_i(\beta_\lambda) + \nabla N_i(\beta_\lambda)(\beta - \beta_\lambda) \quad (18)$$

where

$$\nabla N_i(\beta_\lambda) = \left(\frac{\partial N_i}{\partial \beta_1}, \dots, \frac{\partial N_i}{\partial \beta_r} \right) \Big|_{\beta = \beta_\lambda} \quad (19)$$

Letting $X(\beta_\lambda)$ be the $n \times r$ matrix with ij th entry

$$x_{ij} = \left. \frac{\partial N_i}{\partial \beta_j} \right|_{\beta = \beta_\lambda} \quad (20)$$

and

$$z_i = y_i - N_i(\beta_\lambda) + \sum_{j=1}^n x_{ij} \beta_{\lambda j} \quad (21)$$

we have, that the minimizer of

$$\frac{1}{n} \|z - X(\beta_\lambda) \beta\|^2 + \lambda \beta' \Sigma \beta \quad (22)$$

is, to a good approximation, β_λ . The (approximate) GCV function $V(\lambda)$ is then defined as

$$V(\lambda) = \frac{\frac{1}{n} \text{RSS}(\lambda)}{\left[\frac{1}{n} \text{Tr}(I - A(\lambda, \beta_\lambda)) \right]^2} \quad (23)$$

where

$$\text{RSS}(\lambda) = \sum_{i=1}^n (y_i - N_i(\beta_\lambda))^2 \quad (24)$$

and

$$A(\lambda, \beta_\lambda) = X(\beta_\lambda) (X'(\beta_\lambda) X(\beta_\lambda) + n \lambda I)^{-1} X'(\beta_\lambda). \quad (25)$$

A range of λ -values is explored and the "optimal value" found by minimizing $V(\lambda)$ with respect to λ . In practice it has been found that it is best to do the minimization in a $\log \lambda$ scale. A justification for this procedure may be found in O'S & W.

In O'S & W there were $n = 15$ data points, (= number of channels used for a single column) and a basis of $r = 29$ (>15!) B splines was chosen. The knots of the B-splines were chosen equally spaced in pressure, with multiplicities at the top and bottom. The motivating choice for r was as follows: If r was chosen any larger than 29, it was believed that the estimate of T would not be changed much (that is, minimizing (9) in a collection of functions of the form (8) is just as good as minimizing (9) in the relevant function space), while if r was chosen much smaller, r would act as a smoothing parameter (in addition to the smoothing parameter λ). The following algorithm from O'S & W is probably satisfactory in problems with n and r as large as 100. We shall discuss much larger problems later.

For a given trial value of λ , the minimizer of the objective function (9) is computed via a sequence of Gauss-Newton iterates.

Let $\beta_{\lambda}^{(\ell)}$ be the ℓ th approximate minimizer of (9). At the $(\ell+1)$ st step, $N_i(\beta)$ is expanded about $\beta_{\lambda}^{(\ell)}$ as

$$N_i(\beta) \approx N_i(\beta_{\lambda}^{(\ell)}) + \nabla N_i(\beta_{\lambda}^{(\ell)})(\beta - \beta_{\lambda}^{(\ell)}). \quad (26)$$

Setting

$$x_{ij}^{(\ell)} = \left. \frac{\partial N_i}{\partial \beta_j} \right|_{\beta = \beta_{\lambda}^{(\ell)}} \quad \text{and}$$

$$z_i^{(\ell)} = y_i - N_i(\beta_{\lambda}^{(\ell)}) + \sum_{j=1}^n x_{ij}^{(\ell)} \beta_{\lambda j}^{(\ell)},$$

the iteration is

$$\beta_{\lambda}^{(\ell+1)} = [X^{(\ell)'} X^{(\ell)} + n\lambda \mathbf{I}]^{-1} X^{(\ell)'} z^{(\ell)}, \quad \ell = 1, 2, \dots,$$

where

$$X^{(\ell)} = X(\beta_{\lambda}^{(\ell)}).$$

Once the iteration converges, say at stage L , the GCV function given by (23) with $\beta_{\lambda} = \beta_{\lambda}^L$ is computed, and λ selected as the minimizer of

V. Dropping the superscript (ℓ) , the core of the algorithm can be conveniently implemented using Cholesky factorizations:

1. Find the Cholesky factorization of $X'X + n\lambda \mathbf{I}$,

$$RR' = [X'X + n\lambda \mathbf{I}]$$

2. Solve for β_{λ} by back substitution,

$$RR' \beta_{\lambda} = X'z$$

3. When convergence is reached compute the GCV function

$$V(\lambda) = \frac{\frac{1}{n} \text{RSS}(\lambda)}{[1 + \frac{1}{n} \text{Tr}((RR')^{-1} X'X)]^2}$$

The "forward" calculations of $N_i(\beta_\lambda^{(\ell)})$ and $\left. \frac{\partial N_i}{\partial \beta_j} \right|_{\beta=\beta_\ell}$ were done

using a 40 point quadrature formula to avoid introducing further errors due the numerical analysis. It is known that the solution of ill posed problems can be sensitive to the quality of the quadrature used and we feel it is important not to introduce unnecessary errors at this stage.

4. THREE DIMENSIONAL RETRIEVALS

4.1 EXTENSION OF O'S & W TO THREE DIMENSIONS

Figure 1 presents a schematic diagram of the footprint of TIROS-N, adapted from Susskind et al (1983). Shaded spots represent a hypothetical collection of clear column radiances. Only those clear column radiances will be used to get a three dimensional analysis of temperature in the method proposed below. It is assumed that clear columns can be identified in a separate process. Let $P = (\text{lat.}, \text{long.})$ and p be a vertical coordinate. First, we will suppose that any atmospheric temperature distribution T can be well approximated, for all (P, p) in some volume Ω of the atmosphere by

$$T(P, p) = \sum_{j=1}^r \beta_j B_j(P, p) \quad (27)$$

where the B_j are some appropriately chosen three dimensional basis

functions covering the space within the satellite scan region. We will discuss the choice of the B_j later. The data are modelled as before:

$$y_i = N_i(T) + \epsilon_i, \quad i = 1, 2, \dots, n$$

where now, however, i indexes both the spot and the channel number. If there are S spots with C channels per spot, then there will be $n = SC$ observations. Given a penalty functional $J(T)$ of the form

$$J(T) = \beta' \sum \beta,$$

if $T = \sum_{j=1}^r \beta_j B_j$, the principle in O'S & W is unchanged although in

practice the computational burden will be much greater, and furthermore, approximations and numerical methods appropriate for very large data sets will have to be employed. Accurate feasible quadrature will be required. For given (fixed, or trial) value of λ a numerical method for finding β_λ , the minimizer of

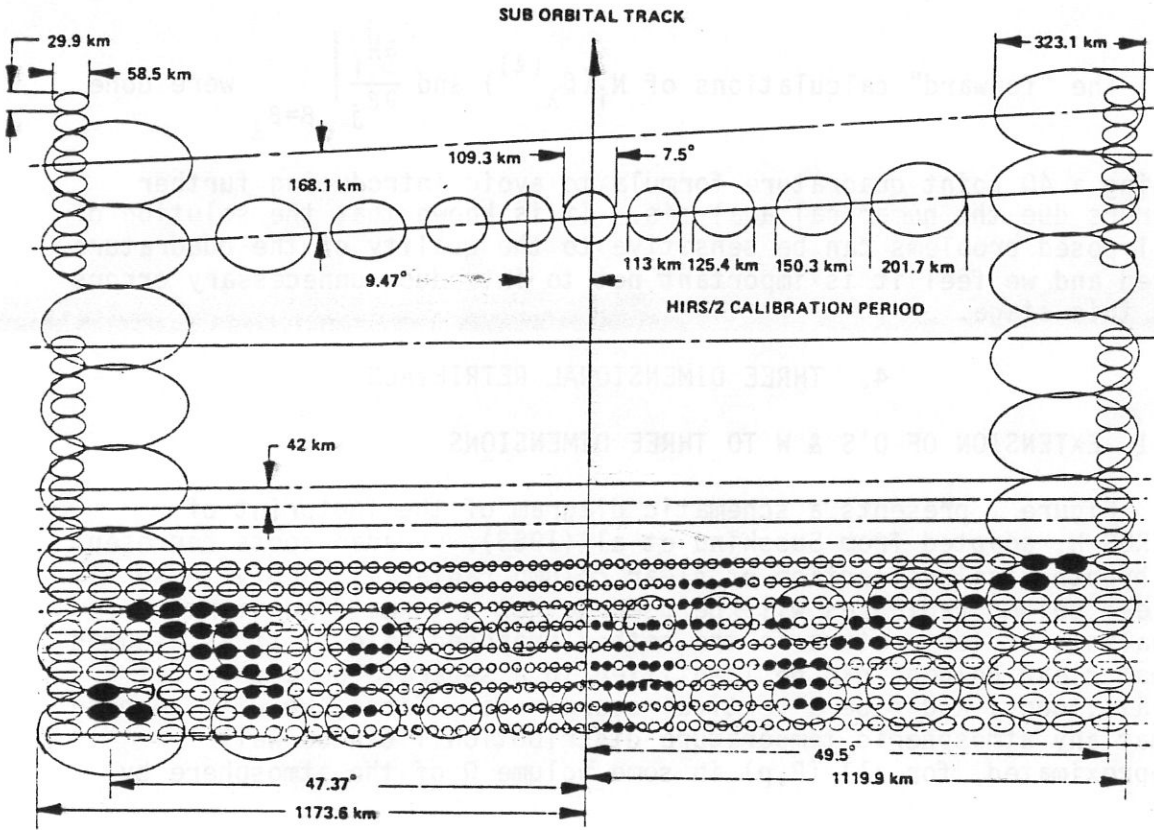


FIGURE 1. TIROS-N footprint

$$\frac{1}{n} \sum_{i=1}^n (y_i - N_i(\beta))^2 + \lambda \beta' \Sigma \beta \quad (28)$$

is required, and if it is required to adaptively tune λ , then it is necessary to evaluate $V(\lambda)$ for several trial values of λ .

Before going on to discuss numerical methods, we will first discuss the choice of the B_j and Σ .

4.2 SPHERICAL HARMONIC EXPANSIONS

Let j be a value of $\{\ell s, k\}$ and let

$$B_{\ell s, k}(P, p) = Y_{\ell s}(P) \phi_k(p) \quad (29)$$

where the $Y_{\ell s}$ are spherical harmonics and the ϕ_k are vertical basis functions, possibly B-splines. There should be enough vertical basis

functions so that any reasonable vertical profile can be adequately approximated by some linear combination of them. A similar observation applies to the number of spherical harmonics. This probably means that there are more vertical basis functions at this stage than model levels.

If

$$T(P,p) = \sum_{\ell s,k} \beta_{\ell s,k} Y_{\ell s}(P) \phi_k(p), \quad (30)$$

then a penalty functional of the form

$$J(\beta) = \sum_{\ell s,k} \beta_{\ell s,k}^2 / \lambda_{\ell s,k} \quad (31)$$

corresponds to a prior covariance⁺ on T of

$$ET(P,p)T(P',p') = \sum_{\ell s,k} \lambda_{\ell s,k} Y_{\ell s}(P) \phi_k(p) Y_{\ell s}(P') \phi_k(p'). \quad (32)$$

(See e.g. Wahba (1979, 1982d), Kimeldorf and Wahba (1971)).

A convenient choice is $\lambda_{\ell s,k}$

$$\lambda_{\ell s,k} = [\ell \cdot \ell + 1]^{-m} w_k \quad (33)$$

for some appropriately chosen m and vertical weights w_k . There

appears to be some reason to believe that $m = 3$ or 4 is a good choice, see Stanford (1979). In principle, m can be chosen by GCV along with λ , see Wahba and Wendelberger (1980). In Wahba (1982d) it is shown how the $\lambda_{\ell s,k}$ for fixed k can be chosen with the

help of data such as that collected by Stanford (1979) and others. In the (unrealistic) case $w_k = 1$, J(β) of (31) with (33) would

correspond to

$$J(T) = \iint [\Delta^m T(P,p)]^2 dP dp,$$

where the integration is taken over the atmosphere. The parameter m embodies information concerning the rate of decay of (horizontal) energy with wavenumber. This expansion has a major numerical advantage, in that the matrix Σ is now diagonal (more on this later) and the covariance function (32) may be chosen with the help of existing meteorological data. For a global analysis or an analysis over a large part of the sphere, it is to be recommended.

⁺We are assuming here and later that the mean has been subtracted out.

In Wahba (1982d) a horizontal wind field analysis from simulated North American radiosonde network data was successfully performed using spherical harmonic representations of the stream function and velocity potential. For a more local analysis, such as over a volume of space above part of the satellite footprint, an expansion in spherical harmonics is still theoretically o.k. The main practical drawback is that a large number of terms (comparable to that for a global analysis) may still be required to maintain numerical accuracy. In this case sensible temperature estimates can be obtained for P, p in the region of dense data, (i.e. estimates of certain linear combinations of the $\beta_{\ell s, k}$'s are good) but a large

number of possibly meaningless estimates of individual $\beta_{\ell s, k}$'s are computed as an intermediate step. In this case the matrix playing the role of X will have many very small eigenvalues. The methods proposed in Bates and Wahba (1983) may alleviate this problem.

4.3 APPROXIMATION BY THIN PLATE SPLINE BASIS FUNCTIONS

The thin plate spline basis functions (TPBF) (Wahba (1980b)) are appropriate when the penalty functional associated with thin plate splines (Wahba and Wendelberger (1980)) is used and data sets are very large. This penalty functional may be suitable when the volume of space in which temperature is to be analyzed is small enough so that the curvature of the earth may be neglected. The TPBF's have been successfully used by Hutchinson and Bischof (1983) and others.

4.4 APPROXIMATIONS BY SECTIONS OF REPRODUCING KERNELS

Suppose that one has a prior covariance in three dimensions

$$ET(P, p)T(P', p') = R(P, p; P' p').$$

One example is the right hand side of (32).

Let $s = (P, p)$ be a point in the atmosphere and let s_1, s_2, \dots, s_r be a regular three dimensional lattice of points in the atmosphere over the region where a temperature analysis is desired. For example a typical s_j is of the form $s_j = (P_{\ell}, p_k)$ where the P_{ℓ} 's are a regular grid in latitude and longitude and the p_k 's are regularly spaced in the vertical. Let us define

$$B_j(P,p) \equiv B_j(s) = R(s,s_j).$$

This notation is intended to mean that s_j is kept fixed and we view R as a function of s . In the approximation theory literature, B_j would be known as a section of the reproducing kernel R - the terminology "reproducing kernel" refers to any positive definite function (i.e. any covariance), but this terminology in addition implies association with a particular Hilbert space of functions (one for each R) known as a reproducing kernel Hilbert space (these spaces are discussed in Aronszajn (1950) see also Kimeldorf and Wahba (1971)). It can be shown that an appropriate penalty functional for this covariance, call it $J_R(f)$ is in fact the square norm of f with the norm of the reproducing kernel space associated with R , and, furthermore, if

$$T = \sum \beta_j B_j,$$

then

$$J_R(T) = \sum_{ij} \beta_i \beta_j R(s_i, s_j). \quad (34)$$

Thus \sum will be the $r \times r$ matrix with ij th entry $R(s_i, s_j)$. (This

formula is related to the terminology "reproducing".) These assertions are easy to check with the specific reproducing kernel given by the right hand side of (32). The reproducing kernel space in question consists of all functions of the form (30) for which (31) is finite. If $T_1 = \sum \beta_{\ell s, k} \gamma_{\ell s} \phi_k$ and $T_2 = \sum \alpha_{\ell s, k} \gamma_{\ell s} \phi_k$, then the inner production this space is

$$\langle T_1, T_2 \rangle = \sum \frac{\beta_{\ell s, k} \alpha_{\ell s, k}}{\lambda_{\ell s, k}}.$$

and $J(T) = \langle T, T \rangle$.⁺ These arguments are of interest in a practical sense since it can be shown (by, for example, using the methods in Wahba (1973)) that any T for which J of (31) is finite can be approximated to a good degree of accuracy in a volume of interest, call it Ω , by some linear combinations of the B_j 's, just provided

there are enough of them with the s_j 's spread around over Ω . We

⁺ $J(T)$ may be modified so that there is no penalty associated with the constant function or a few very low wavenumbers, see Wahba (1979).

remark that, for the penalty functional $J(f) = \int (f^{(m)}(x))^2 dx$ the B-splines of degree $2m-1$ may be obtained as linear combinations of sections of a reproducing kernel and the TPBF's of Section 4.3 can also be shown to be linear combinations of sections of a reproducing kernel. The good approximation-theoretic properties of B-splines are well known. These results give us a tool for approximation. For example, suppose that $\lambda_{\ell s, k}$ in (32) has the special form $\lambda_{\ell s, k} = \lambda_{\ell k}$. Then

$$\begin{aligned} R(P, p; P', p') &= \sum_{\ell s, k} \lambda_{\ell k} Y_{\ell s}(P) Y_{\ell s}(P') \phi_k(p) \phi_k(p') \\ &= \sum_k R_k(P, P') \phi_j(p) \phi_j(p'), \text{ say,} \end{aligned} \quad (35)$$

where

$$R_k(P, P') = \sum_{\ell s} \lambda_{\ell k} Y_{\ell s}(P) Y_{\ell s}(P') \quad (36)$$

Then, for example, for $s_j = (P_\ell, p_k)$ we have

$$B_j(P, p) = \sum_v R_j(P, P_\ell) \phi_j(p) \phi_v(p_k). \quad (37)$$

Note that each B_j has its maximum at $(P, p) = (P_\ell, p_k) = s_j$ and near $s = s_j$ it behaves like a hill function.

In general, it may be mathematically more convenient to represent this basis set by

$$B_j(P, p) = R_k(P, p_\ell) \phi_k(p). \quad (38)$$

Then, if $j = (\ell, k)$ and $j' = (\ell', k')$, the jj' th entry of Σ is 0 if $k \neq k'$ and $R_k(P_\ell, P_{\ell'})$ if $k = k'$. Some closed form formulae for $R_k(P, P')$ which approximate $R(P, P')$ with $\lambda_{\ell k} = [\ell(\ell+1)]^{-m}$

may be found in Wahba (1981b, 1982b). These basis functions appear to be well suited for the purpose of extrapolating over "holes" in the data in a way that uses the information in the prior covariance in a natural way. Their major drawback is that they do not have local support, which means that computing with them is relatively expensive. In one dimension the local support B-splines may be obtained as linear combinations of a reproducing kernel (associated

with $\int (f^{(m)}(x))^2 dx$) but it does not appear that this is possible (exactly) in more than one dimension). Some approximations will have

to be made to obtain local support bases well suited to the penalty functionals considered here. Dierckx (1983) has recently proposed tensor products of B-splines as basis functions on the sphere. Further research in connection with the choice of basis functions is clearly warranted.

5. SIMULTANEOUS ANALYSIS OF FORECAST, RADIOSONDE AND RADIANCE INFORMATION

5.1 RADIOSONDE DATA

Suppose radiosonde data at points $t = (P, p)$, $t = t_1, \dots, t_N$ are available in the volume of space over which temperature is to be analyzed. Letting the B_j 's be as in Section 4, the radiance data is modelled as

$$\begin{aligned} y_\ell &= T(t_\ell) + \epsilon_\ell, \quad \ell = 1, 2, \dots, n' \\ &\approx \sum_{j=1}^{n'} \beta_j B_j(t_\ell) + \epsilon_\ell \end{aligned} \quad (39)$$

and the variational problem of (28) can be modified to

$$\sum_{i=1}^n (y_i - N_i(\beta))^2 + w_R \sum_{\ell=1}^{n'} (y_\ell - T_\ell(\beta))^2 + \lambda \beta' \Sigma \beta \quad (40)$$

where $T_\ell(\beta) = \sum_{j=1}^{n'} \beta_j B_j(t_\ell)$, scaling factors similar to the w_i of (6)

which reflect relative accuracy in the temperature measurements have been absorbed into the notation and w_R is a tuning parameter governing

the relative weight to be given to radiance and radiosonde information. In theory one should be able to specify w_R a priori but

in practice unknown variations in the accuracy in measuring (and modelling) the radiances probably make it appropriate to leave w_R as a tuning parameter.

5.2 FORECAST DATA

Suppose one has a spectral model expanded in terms of the B_j 's. (This is conceivable if spherical harmonics are used.) Then one has "forecasts" β_j^F of the β_j 's. In this case one can add a term of the form

$$w_F \sum_j q_{jj} (\beta_j^F - \beta_j) (\beta_j^F - \beta_j) \quad (41)$$

to (40). Here the q_{jj} are based on some (multiple of) the presumed covariance matrix of the β_j^F 's compared to the "true" β_j 's. In this case q_{jj} may be 0 for $j \neq j'$, which would introduce a simplification.

See Wahba (1982c) for a justification of this procedure. We believe that an automatic tuning procedure for choosing w_R and w_F based on

a GCV-like statistic can be found (work in progress). The argument here and in Wahba (1982c) leads us to believe that available heterogeneous data sets may be combined in a single variational problem along with forecasts of the spectral coefficients in a spectral model to update the coefficients in a spectral model.

6. TROPOPAUSE HEIGHT INFORMATION

Suppose tropopause height is known on, say a regular grid P_1, \dots, P_L in P . If tropopause height is observed on an irregular grid e.g. by the radiosonde network an analysis to arbitrary points may be obtained e.g. by using either the TPS of Wahba and Wendelberger (1980) or the splines on the sphere of Wahba (1981b), that is, the approach described here reduced to the two dimensional, linear case. Consider the following side condition:

$$\left. \frac{\partial T}{\partial p} \right|_{p \rightarrow \text{trop} - \delta} \leq 0 \quad P = P_1, \dots, P_L \quad (42)$$

$$\left. \frac{\partial T}{\partial p} \right|_{p \rightarrow \text{trop} + \delta} \geq 0 \quad P = P_1, \dots, P_L. \quad (43)$$

If $T(P_\ell, p) = \sum_j \beta_j B_j(P_\ell, p)$, the inequality (42) becomes

$$\left. \sum_j \beta_j \frac{\partial B_j}{\partial p}(P_\ell, p) \right|_{p \rightarrow \text{trop} - \delta} \leq 0 \quad \ell = 1, 2, \dots, L \quad (44)$$

which is a set of linear inequality constraints on the β_j . Similarly

for (43). In principle the variational problem (40), with the possible inclusion of the term (41) can be solved subject to these linear inequality constraints (provided there are sufficient

vertical basis functions). For a good answer, there should be sufficiently many vertical basis functions so that the minimum at the tropopause can be described without undue distortion to the remainder of the vertical profiles. In practice Villalobos (1983) and Villalobos and Wahba (in preparation) have solved such problems with linear data functionals (N_i 's) and several hundred data points and

inequality constraints, using a form of GCV adapted to linear inequality constraints, and the algorithm of Gill et al (1984). We believe that numerical methods may be developed to handle larger data sets and numbers of linear inequality constraints.

7. SIMULTANEOUS ANALYSIS OF TEMPERATURE AND WATER VAPOR

Bill Smith (this volume, and personal communication) has proposed a simultaneous analysis of water vapor and temperature. Some of his ideas can be incorporated into the present approach as follows. Let $U(P,p)$ be the water vapor and suppose the water vapor is expanded in some basis functions $\{B_j\}$ which may or may not be the same basis

functions as used for temperature:

$$U(P,p) = \sum_j \alpha_j B_j(P,p). \quad (45)$$

The dependence of transmittance τ_v of Section 2 on water vapor is now explicitly modelled, and $N_i(T) = N_i(\beta)$ of Sections 2 and 3 is now replaced by $N_i(T,U) = N_i(\beta, \alpha)$.

The analysis problem of (9) now becomes, in its simplest form: Minimize

$$\sum_{i=1}^n (y_i - N_i(\beta, \alpha))^2 + \lambda_\beta \beta' \sum \beta_\beta \beta + \lambda_\alpha \alpha' \sum \alpha_\alpha \alpha. \quad (46)$$

In (46) known relations between U and T are being ignored, they may be inserted by replacing the simplified penalty functional of (46) by a more general suitable quadratic form involving cross terms in α and β and/or by including linear inequality relations known to hold between temperature and water vapor content. If radiosonde data of the form

$$z_\ell = U_\ell(\alpha) + \epsilon_\ell,$$

is available, where $U_\ell(\alpha) \approx \sum_j \alpha_j B_j(t_\ell)$, then a term of the form

$$w_{R,U} \sum_{\ell=1}^{n'} (z_\ell - U_\ell(\alpha))^2$$

may be added to (46), similarly if forecast information is available a term of the form analogous to (41) may be added to (46), along with the analogous temperature terms. For example one possibility is: minimize

$$\begin{aligned} & \sum (y_i - N_i(\beta, \alpha))^2 + w_{R,T} \sum (y_\ell - T_\ell(\beta))^2 + w_{R,U} \sum (z_\ell - U_\ell(\alpha))^2 \\ & + w_{F,T} \sum q_{jj',T} (\beta_j^F - \beta_{j'}) (\beta_{j'}^F - \beta_j) \\ & + w_{F,U} \sum q_{jj',U} (\alpha_j^F - \alpha_{j'}) (\alpha_{j'}^F - \alpha_j) \\ & + \lambda J(\beta, \alpha) \end{aligned} \quad (47)$$

where J is a suitable penalty functional in β and α which may contain one or more additional tuning parameters.

8. NUMERICAL METHODS FOR LARGE, CONSTRAINED VARIATIONAL PROBLEMS

In its most general form the variational problem to be solved is of the form: find $\gamma = (\gamma_1, \dots, \gamma_R)'$ to minimize

$$\left\{ \sum_{i,j=1}^N (y_i - L_i(\gamma)) q_{ij}(w) (y_j - L_j(\gamma)) \right\} + \lambda J(\gamma, \theta) \quad (48)$$

(possibly) subject to the linear inequality constraints

$$\sum_j c_{ij} \gamma_j \geq \delta_i \quad i = 1, 2, \dots, N'. \quad (49)$$

where y_i , $i = 1, \dots, N$ are "observations" (which includes scaled observations as well as forecasts), L_i is a linear or nonlinear functional of the unknown "state" vector γ , the $q_{ij}(w)$ represent a positive definite quadratic form which may depend on one or more weights (w) to be chosen, and $J(\gamma, \theta)$ is a nonnegative definite quadratic form in γ which may further depend on some tuning parameters θ . Similar variational problems are proposed in Hoffman (this volume). If the L_i are linear and the term in brackets has a

unique minimizer in the null space of J (or, if J is strictly positive definite) then (48) will have a unique minimizer for all y , and it will also have a unique minimizer among all γ satisfying the constraints (49). Some conditions for the existence of a minimizer when the L_i are nonlinear are given in O'Sullivan (1983).

GCV may be used to find a good value of λ and (usually) θ in the constrained, not necessarily quadratic problem defined by (48) and (49), by combining the GCV for nonlinear data functionals in O'S and W with the GCV for constrained problems in Wahba (1980, 1982e) and Villalobos (1983). The general idea is as follows: The solution γ_λ to the variational problem of (48) and (49) is found for a trial value of λ . Then an equality-constrained quadratic problem approximating the problem of (48) and (49) at γ_λ is found. (A change of variables should be made so that the quadratic form approximating the term in brackets in (48) is a sum of squares). Then the $A(\lambda)$ matrix for this problem is used to compute $V(\lambda)$.

In general, the numerator $RSS(\lambda)$ will usually be available once γ_λ is found. It may be a major expense to compute

$$\frac{1}{N} \text{Tr} A(\lambda) = \frac{1}{N} \sum_{i=1}^N a_{ii}(\lambda). \quad \text{A relatively cheap shortcut using}$$

bidiagonalization appears in Elden (1983). The $a_{ii}(\lambda)$ are all between

0 and 1. An estimate of $\frac{1}{N} \sum_{i=1}^N a_{ii}(\lambda)$ may be made, if N is very large

by computing a random or a stratified sample of the $a_{ii}(\lambda)$ and averaging.

If w is fixed, and the L_i linear the unconstrained problem (48)

with GCV may be computed for R of the order of a few hundred by the methods in Bates and Wahba (1983). It is probably possible to combine the Gauss-Newton iteration in O'S and W (1984) with those methods to solve the unconstrained nonlinear problem. Hoffman (1984) and Testud and Chong (1983) have solved problems similar to the unconstrained problem with N of the order to several thousand and R of the order of 1,000 with sparse matrices by conjugate gradient methods. Herman, Lent and Hurwitz (1980) provide a storage efficient method of minimizing $\|y - X\gamma\|^2 + \lambda \sum_j \gamma_j^2$ which they

claim has been used with $N \sim 200,000$ and $R \sim 600,000$. Iterative row action methods for solving large linear systems like those used by Fleming (1983), see Censor (1981) may possibly provide the core of an optimization algorithm suitable for (48) and (49) with very large N and R . As Fleming has observed, stopping the iteration early in these methods is a form of regularization, and it may be possible to exploit this fact. (See Wahba (1980a, Section 7). In general, sequential quadratic programming methods, e.g. as discussed by Gill, Murray and Wright (1981) are appropriate for the constrained optimization problem, particularly when the GCV function is to be

computed. Although we do not know just how large a problem can be handled by those methods it is quite likely that R of 1,000 or more is feasible with parallel processing. It is observed that these problems with non sparse matrices are quite suitable for parallel processing.

We remark that under some circumstances, techniques for outlier detection and confidence intervals are available in conjunction with these variational methods, but we omit the details. (See, for example Wahba (1983b), Eubank (1984).

Clearly further research on numerical methods for these problems is needed but we believe it will bear fruit.

We would like to thank Don Johnson, Bill Smith, Walter Murray and Doug Bates for helpful conversations and Ross Hoffman for providing us early drafts of his work.

REFERENCES

- Aronszajn, N., (1950), Theory of reproducing kernels, Trans. Am. Math. Soc., 68, 337-404.
- Bates, D. and Wahba, G., (1983), A truncated singular value decomposition and other methods for generalized cross-validation, University of Wisconsin-Madison Statistics Dept. TR 715.
- Censor, Y., (1981), Row-action methods for huge and sparse systems and their applications, SIAM Review, 23, 444-466.
- 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.
- DeBoor, C., (1978), A Practical Guide to Splines, New York: Springer-Verlag.
- Dierckx, P., (1984), Algorithms for smoothing data on the sphere with tensor product splines, Computing, 32, 319-342.
- Elden, L., (1983), A note on the computation of the generalized cross-validation function for ill-conditioned least squares problems, Department of Mathematics, Linköping University, Report LiTH-MAT-R-1983-27, to appear, BIT, 24, 4.
- Eubank, R., (1984), The hat matrix for smoothing splines, Statistics and Probability Letters, 2, 9-14.
- Fleming, H., (1983), Satellite remote sensing by the technique of computed tomography, J. Applied Meteorology, 21, 1539-1549.

- Fleming, H.E. and Crosby, D.S., (1983), Effect of incorrect atmospheric transmittance on minimum variance satellite temperature retrievals, in Preprint Volume, Eighth Conference on Probability and Statistics in Atmospheric Sciences, Boston, MA: American Meteorological Society 73-76.
- Fleming, H.E., Crosby, D.S., and Neuendorffer, A.C., (1983), Elimination of a major error component in satellite temperature soundings, in Preprint Volume, Fifth Conference on Atmospheric Radiation, Boston, MA: American Meteorological Society 20-22.
- Fritz, S., Wark, D., Fleming, J., Smith, W., Jacobowitz, H., Hilleary, D., and Alishouse, J. (1972), Temperature sounding from satellites, NOAA Tech. Rept. NESS 59, National Oceanic and Atmospheric Administration, Washington, D.C.
- Gill, P., Gould, N., Murray, W., Saunders, M., and Wright, M., (1984), A weighted Gram-Schmidt method for convex quadratic programming, Mathematical Programming, 30, 176-194.
- Gill, P., Murray, W., and Wright, M., (1981), Practical Optimization, New York: Academic Press.
- Herman, G. T., Lent, A., and Hurwitz, H., (1980), A storage-efficient algorithm for finding the regularized solution of a large, inconsistent system of equations, J. Inst. Maths. Applies, 25, 361-366.
- Hoffman, R. N. (1983), Three dimensional inversion of satellite observed radiances: A proposal, Preprint Volume: Fifth Conference on Atmospheric Radiation, American Meteorological Society, 43-46.
- Hoffman, R.N., (1984), SASS Wind Ambiguity Removal by Direct Minimization. Part II: Use of Smoothness and Dynamical Constraints, Mon. Wea. Rev. (submitted), Atmospheric and Environmental Research, Inc., (manuscript).
- Hutchinson, M. and Bischof, R., (1983), A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales, Aust. Met. Mag., 31, 179-184.
- Kimeldorf, G. and Wahba, G., (1971), Some results on Tchebycheffian spline functions, J. Math. Anal. Applic., 33, 82-95.
- O'Sullivan, F., (1983), The analysis of some penalized likelihood estimation schemes, University of Wisconsin-Madison, Statistics Department Technical Report No. 726.

- O'Sullivan, F. and Wahba, G., (1984), A cross validated Bayesian retrieval algorithm for non-linear remote sensing experiments, J. Comput. Physics, to appear. University of Wisconsin-Madison Statistics Department Technical Report No. 747.
- Smith, W., (1983), The retrieval of atmospheric profiles from VAS geostationary radiance observations, J. Atmospheric Sciences, 40, 2025-2035.
- Smith, W. (1984), Simultaneous retrieval of surface and atmospheric parameters - A physically and analytically direct solution, this volume.
- Stanford, J., (1979), Latitudinal-wavenumber power spectra of stratospheric temperature fluctuations, J. Atmospheric Sciences, 36, 921-931.
- Susskind, J., Rosenfield, J., and Reuter, D., (1983), An accurate radiative transfer model for use in the direct physical inversion of HIRS2 and MSU temperature sounding data, J. Geophysical Research, 88, 8550-8568.
- Susskind, J., Rosenfield, J., Reuter, D., and Chahine, M. T., (1984), Remote sensing of weather and climate parameters from HIRS/MSU on TIROS-N, J. Geophysical Research, 89, 4677-4697.
- Testud, J. and Chong, M., (1983), Three-dimensional wind field analysis from dual-doppler radar data. Part I: Filtering, interpolating and differentiating the raw data, J. Climate and Applied Meteorology, 22, 1204-1215.
- Villalobos, M., (1983), Estimation of posterior probabilities using multivariate smoothing splines and generalized cross-validation, Department of Statistics, University of Wisconsin-Madison, TR No. 725. (thesis)
- Wahba, G., (1973), Convergence rates of certain approximate solutions to Fredholm integral equations of the first kind, J. Approx. Theory, 7, 167-185.
- Wahba, G., (1977), Practical approximate solutions to linear operator equations when the data are noisy, SIAM J. Numerical Analysis, 14, 651-667.
- Wahba, G., (1979), Spline interpolation and smoothing on the sphere, Statistics Dept., University of Wisconsin-Madison, Technical Report No. 584.

- Wahba, G., (1980a), Ill posed problems: Numerical and statistical methods for mildly, moderately, and severely ill posed problems with noisy data. University of Wisconsin-Madison Statistics Department Technical Report No. 595. (To appear in the Proceedings of the International Conference on Ill Posed Problems, M.Z. Nashed, ed.)
- Wahba, G., (1980b), Spline bases, regularization, and generalized cross validation for solving approximation problems with large quantities of noisy data, in Approximation Theory III, ed. W. Cheney, Academic Press 905-912.
- Wahba, G., (1981a), Cross validation and constrained regularization methods for mildly ill posed problems. University of Wisconsin-Madison, Statistics Department Technical Report #629.
- Wahba, G., (1981b), Spline interpolation and smoothing on the sphere, SIAM J. Sci. Stat. Comput., 2, 5-16.
- Wahba, G., (1982a), Constrained regularization for ill posed linear operator equations, with applications in meteorology and medicine, in Statistical Decision Theory and Related Topics III, Vol. 2, ed. S.S. Gupta and J.O. Berger, Academic Press 383-418.
- Wahba, G. (1982b), Erratum: Spline interpolation and smoothing on the sphere, SIAM J. Sci. Stat. Comput., 3, 385-386.
- Wahba, G. (1982c), Variational methods in simultaneous optimum interpolation and initialization, in Proceedings of the (Fourteenth) Stanstead Seminar, ed. D. Williamson, NCAR and McGill University 178-185.
- Wahba, G., (1982d), Vector splines on the sphere, with application to the estimation of vorticity and divergence from discrete, noisy data, in Multivariate Approximation Theory, Vol. 2, ed. W. Schempp and K. Zeller, Birkhauser Verlag 407-429.
- Wahba, G., (1982e), 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. Berger, eds., Academic Press, 383-418.
- Wahba, G., (1983a), Eigensequence plots and design criteria for satellite computed tomography, Dept. of Statistics, University of Wisconsin-Madison TR No. 732, (submitted).
- Wahba, G., (1983b), Bayesian "confidence intervals" for the cross-validated smoothing spline, J. Roy. Stat. Soc. B, 45, 133-150.
- Wahba, G. and Wendelberger, J., (1980), Some new mathematical methods for variational objective analysis using splines and cross-validation, Monthly Weather Review, 108, 36-57.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 755	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) VARIATIONAL METHODS FOR MULTIDIMENSIONAL INVERSE PROBLEMS		5. TYPE OF REPORT & PERIOD COVERED Scientific Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Grace Wahba		8. CONTRACT OR GRANT NUMBER(s) NASA NAG5-316 ONR N00014-77-C-0675
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of Wisconsin, 1210 W. Dayton St. Madison, WI 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 N. Quincy St. Arlington, VA 22217		12. REPORT DATE December 1984
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) multidimensional inverse problems remote sensing illposed problems three dimensional satellite temperature retrievals		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) (see reverse side)		

ABSTRACT

By combining ideas from various sources, we propose a class of physical variational methods for estimating three dimensional temperature structure from satellite radiance data using many sets of vertical and/or oblique soundings simultaneously. The class of methods proposed takes advantage of assumed slowly varying temperature structure in the horizontal and linearizes the inherently (mildly) nonlinear problem of temperature retrievals as late as possible. The method can handle irregularly spaced clear column radiances. Methods are proposed for the specific inclusion and weighting of forecast information, radiosonde data and tropopause height information. Following Smith (1984) it is suggested how temperature and water vapor may be simultaneously analyzed with the proposed methods. So long as nature is "smooth" or highly correlated it is in principle advantageous to build this "smoothness" or correlation information into the analysis, via analyzing large quantities of data simultaneously. A method for using GCV (generalized cross validation) to get (some of) the tuning parameters adaptively in nonlinear and constrained problems with large data sets is described. The main drawback to dealing with large sets of data simultaneously is the computational cost in time and storage. Various algorithms and shortcuts are proposed for approaching the computational problem.