DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT NO. 747

July 1984

A CROSS VALIDATED BAYESIAN RETRIEVAL ALGORITHM FOR NON-LINEAR REMOTE SENSING EXPERIMENTS

by

Finbarr O'Sullivan and Grace Wahba

1980 Mathematics Subject Classification

65R20 Integral equations
65K10 Optimization and variational techniques
65D10 Smoothing, curve fitting
65D07 Splines
86A10 Meteorology

Research of F. O'Sullivan sponsored by NSF under Grant MCS-8403239

Research of G. Wahba sponsored by ONR under Contract N00014-77-C-0675 and NASA under Grant NAG5-316

A CROSS VALIDATED BAYESIAN RETRIEVAL ALGORITHM FOR NON-LINEAR REMOTE SENSING EXPERIMENTS

Finbarr O'Sullivan

Department of Statistics
University of California
Berkeley. CA 97420.

Grace Wahba

Department of Statistics
University of Wisconsin
Madison. WI 53706.

ABSTRACT

We present a fully automated retrieval algorithm for estimating non-linearly *sensed* functions. This algorithm combines a simple Gauss-Newton iteration with an extended form of Generalized Cross Validation. The performance of the algorithm is illustrated in the context of a remote sensing problem arising in satellite meteorology.

1. Introduction

Remote sensing experiments may be found in every branch of the physical and biological sciences. In these experiments, the desired phenomenon cannot be observed directly, rather remote or indirect observations are available on some functionals of the function of interest. Some examples are: computerized tomography, recovery of geological structure via seismic exploration, the recovery of atmospheric vertical temperature structure from satellite-observed upwelling radiation, and the recovery of particle size distributions from scattered radiation or cross-sectional information. For these experiments the usual observational model is:

$$y_i = N(\theta, x_i) + \epsilon_i, \qquad i = 1,2, \ldots, n. \tag{1}$$

$\theta(t)$, for t in Ω, is the unknown *smooth* function it is desired to recover, the ϵ_i's are zero-mean measurement errors with variance $E\epsilon_i^2 = w_i^{-1}\sigma^2$, where the scale factor σ^2 may be unknown. Ω could be the real line, Euclidean d-space, the sphere or the atmosphere. The x_i's are

the design points, and $N(\theta, x_i)$ is, for each i, a non-linear functional of θ. For example

$$N(\theta, x_i) = \int K(x_i, \theta(t), t)\, dt. \tag{2}$$

The design space, (the space containing the x_i's,) may be quite arbitary.

It is desired to retrieve an estimate of the function θ from the observed data. A wealth of procedures have been proposed for handling such problems. See Backus and Gilbert[2], Chahine[4], Smith[21], Surmont and Chen[24], and also the books of Anderssen, de Hoog and Lukas[1], Baker and Miller[3], Deepak[9], Groetsch[11], Rall[20], Tikhonov and Arsenin[25] and Twomey[26]. Recently, Bayesian or Regularization techniques, and especially the cross-validated spline smoothing methods proposed in Craven and Wahba[6], Golub Heath Wahba[10] and Wahba[27] have had some popular success in this area, see Crump[7], Hutchinson and Bischof[12], Merz[16] and Nychka *et al*[17]. In a Bayesian framework, the estimation of a function is done by choosing a value of the function which is probable in the light of the observed data, but is also in keeping with prior notions about the behavior of that function. For the remote-sensing model this reduces to estimating θ by minimizing a quantity of the form

$$I_\lambda(\theta) = \frac{1}{n}\sum_{i=1}^{n} w_i[y_i - N(\theta, x_i)]^2 + \lambda J(\theta) \qquad \lambda > 0 \tag{3}$$

over some plausible set of candidates Θ. The first term measures fidelity to the observed data, while the second term, $J(\theta)$, takes the prior information into account. Here we let J be a quadratic functional of θ which is zero when θ is the zero function; (e.g. as $J(\theta) = \int [\theta''(t)]^2 dt$). The smoothing parameter, λ, controls the relative weighting given to the prior. Formally, λ is a free parameter to be chosen by the analyst, but in practical situations it is convenient to have a reliable automatic procedure for isolating a ball-park value. In this paper we describe how the Generalized Cross-Validation of Craven and Wahba[6], can be adapted to handle this problem. The result is a fully automatic retrieval algorithm for non-linear remote sensing experiments.

Our analysis could have been carried out in a function space setting, however, we decided to avoid this generality. Here we suppose that the function θ can be well approximated by a

finite collection of K basis functions ϕ_ν *i.e.*

$$\theta \approx \sum_{\nu=1}^{K} \beta_\nu \phi_\nu \tag{4}$$

where $\beta = (\beta_1, \cdots, \beta_K)'$ is an element of R^K. If n is small, or even of moderate size, K will generally be larger than n, we discuss the choice of K in a particular example later. With this, the remote sensing problem is rephrased as the estimation of the coefficients β from the data. Substituting into Eq.(3) and absorbing w_i inside the brackets, the estimate of β minimizes:

$$\frac{1}{n}\sum_{i=1}^{n}[z_i - \eta(\beta, x_i)]^2 + \lambda \beta' \Sigma \beta \tag{5}$$

where $z_i = \sqrt{w_i} y_i$, $\eta(\beta, x_i) = \sqrt{w_i} N[\sum_{\nu=1}^{K} \beta_\nu \phi_\nu, x_i]$ and, since J is quadratic, $\beta' \Sigma \beta = J(\sum_{\nu=1}^{K} \beta_\nu \phi_\nu)$. In a practical problem the approximation properties of the basis functions should be carefully considered in the light of the resolving power of the instrument and, of course, the measurement noise. In this connection, B-spline basis functions, described in the book by De Boor[8], have three very attractive features:

(i) Excellent approximation properties.

(ii) Local support which, in integral equation settings, can simplify numerical quadrature.

(iii) Publically available software for computational manipulation.

From a theoretical point of view the introduction of the basis functions is not necessary, it will, we hope, make the algoritm a bit more transparent. A discussion of the more elaborate function space theory can be found in O'Sullivan[18] and the references cited therein. The paper is organized as follows: The retrieval algorithm is presented in Section 2. The algorithm selects an appropriate value for the smoothing parameter by minimizing an extended Generalized Cross Validation (GCV) function. In Section 3 we review the idea of GCV for linear problems and motivate the extension to the non-linear case. Section 4 talks about a temperature profile retrieval problem which arises in Satellite Meteorology and reports on a small

Monte Carlo study illustrating the performance of the method.

2. The Automatic Bayesian Retrieval Algorithm

For a given trial value of λ, the minimizer of the objective function, Eq.(5), is computed via a sequence of Gauss-Newton iterates. Let $\beta_\lambda^{(l)}$ be the l'th approximate minimizer of (5). At the $(l+1)$'st step, $\eta(\beta,x_i)$ is expanded about $\beta_\lambda^{(l)}$ as:

$$\eta(\beta,x_i) \approx \eta(\beta_\lambda^{(l)},x_i) + \nabla\eta(\beta_\lambda^{(l)},x_i)(\beta - \beta_\lambda^{(l)}) \tag{6}$$

and the new iterate is defined to be the minimizer of the quadratic function

$$\sum_{i=1}^{n}[z_i - \eta(\beta_\lambda^{(l)},x_i) - \nabla\eta(\beta_\lambda^{(l)},x_i)(\beta - \beta_\lambda^{(l)})]^2 + \lambda\beta'\Sigma\beta. \tag{7}$$

So letting $X_{ij}^{(l)} = \dfrac{\partial\eta(\beta,x_i)}{\partial\beta_j}\Big|_{\beta=\beta_\lambda^{(l)}}$ and $z_i^{(l)} = X_i^{(l)}\beta_\lambda^{(l)} + z_i - \eta(\beta_\lambda^{(l)},x_i)$ the iteration is

$$\beta_\lambda^{(l+1)} = [X^{(l)\prime}X^{(l)} + n\lambda\Sigma]^{-1}X^{(l)\prime}\underline{z}^{(l)} \qquad l = 1,2... \tag{8}$$

Once the iteration numerically converges, say at stage L, λ is assessed by the extended GCV function, $V(\lambda)$, described in section 3 below.

$$V(\lambda) = \frac{\frac{1}{n}RSS(\lambda)}{[tr\frac{1}{n}(I-A_L(\lambda))]^2}. \tag{9}$$

where $RSS(\lambda) = \sum_{i=1}^{n}[z_i-\eta(\beta_\lambda,x_i)]^2$ is the residual sum of squares and $A_L(\lambda) = X^{(L)}[X^{(L)\prime}X^{(L)} + n\lambda\Sigma]^{-1}X^{(L)\prime}$. A range of λ-values is explored and the 'optimal' value found by minimizing $V(\lambda)$ *w.r.t.* λ. In practice, it has been found that it is best to do this minimization in log λ scale. Dropping superscripts the core of the algorithm can be conveniently implemented using Cholesky factorizations:

1. Find the Cholesky factorization of $X'X + n\lambda\Sigma$

$$RR' = [X'X + n\lambda\Sigma]$$

2. Solve for β_λ by back substitution

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

3. When convergence is reached compute the GCV

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

The calculation of the X-matrices will necessitate integration of the basis functions whenever the functionals $N(\cdot,x_i)$ involve integrals of θ. Numerically such integrations can be rapidly performed if the basis functions have local support (such as B-splines). Aside from the computation of the iterative design matrices, the operation count of the algorithm is dominated by the $O(K^3)$ Cholesky steps. This is to be expected since the technique ignores any structure which the matrices $[X^{(l)'}X^{(l)} + n\lambda\Sigma]$ may possess. In particular situations it would be desirable to exploit any structure which these matrices possessed.

3. Generalized Cross Validation

In our retrieval algorithm values of the smoothing parameter are assessed by means of a Generalized Cross Validation function. Generalized Cross Validation (GCV) was developed by Craven and Wahba[6], for assessing λ in the case where the data functionals were linear. Here we describe how this technique works and provide an extension to the non-linear case.

3.1. The Linear Case

If $\eta(\beta,x_i)$ is linear then we may write $\eta(\beta,x_i) = X_i\beta$ where X_i is an element of R^K. It follows that the estimate, β_λ, solves the linear system

$$[X'X + n\lambda\Sigma]\beta_\lambda = X'z \tag{10}$$

where the rows of X are given by the vectors $X_i = (X_{i1}, X_{i2}, \cdots, X_{iK})$. Assuming $[X'X + n\lambda\Sigma]$ is of full rank,

$$\beta_\lambda = [X'X + n\lambda\Sigma]^{-1}X'z, \tag{11}$$

and the predictions are

$$\begin{aligned} X\beta_\lambda &= X[X'X + n\lambda\Sigma]^{-1}X'z \\ &\equiv A(\lambda)z \quad , \text{ say.} \end{aligned} \tag{12}$$

In statistical jargon, A is the hat-matrix for the regression problem. With this notation, Craven and Wahba's GCV function may be written as

$$V(\lambda) = \frac{\frac{1}{n}\sum_{i=1}^{n}[z_i - X_i\beta_\lambda]^2}{[tr\frac{1}{n}(I - A(\lambda))]^2}$$

$$\equiv \frac{\frac{1}{n}RSS(\lambda)}{tr[\frac{1}{n}(I - A(\lambda))]^2}. \tag{13}$$

The $tr(I - A(\lambda))$ can be viewed as an estimate of degrees of freedom for error. As $\lambda \to 0$ the residual sum of squares and the degrees of freedom for error are made small, while as $\lambda \to \infty$ both these terms become large. Consequently, $V(\lambda)$ is an assessment of λ which balances degrees of freedom for error against model fit. More formally, Golub, Heath and Wahba[10], motivate $V(\lambda)$ as a rotational invariant version of *ordinary* cross validation. The ordinary cross validation function, $V_0(\lambda)$, is defined as

$$V_0(\lambda) = \sum_{k=1}^{n}[z_k - X_k\beta_\lambda^{[k]}]^2 \tag{14}$$

where $\beta_\lambda^{[k]}$ is the minimizer of the objective function with the k'th data point omitted, *i.e.* $\beta_\lambda^{[k]}$ minimizes

$$\frac{1}{n}\sum_{i\neq k}[z_i - X_i\beta]^2 + \lambda\beta'\Sigma\beta. \tag{15}$$

The idea is that if λ is a good choice, then $X_k\beta_\lambda^{[k]}$ should be, on average, a good predictor of z_k. Let $\beta[u_k]$ be the solution in Eq.(10) with the k'th data point, z_k, replaced by u_k. Then working from the identity, proven in Lemma 3.1 of[6],

$$X_k\beta_\lambda[z_k + \delta_k] = X_k\beta_\lambda^{[k]} \tag{16}$$

where $\delta_k = X_k\beta_\lambda^{[k]} - z_k$. One can verify, by direct algebra, that

$$V_0(\lambda) = \frac{1}{n}\sum_{k=1}^{n}\frac{[z_k - X_k\beta_\lambda]^2}{[1 - a_{kk}(\lambda)]^2} \tag{17}$$

where

$$a_{kk}(\lambda) = \frac{X_k\beta_\lambda[z_k + \delta_k] - X_k\beta_\lambda[z_k]}{\delta_k} \equiv A_{kk}(\lambda),$$

is the k'th diagonal entry of $A(\lambda)$. Finally, if we replace $a_{kk}(\lambda)$ by $\frac{1}{n}\sum_{j=1}^{n} a_{jj}(\lambda)$ then V_0 is rotationally invariant, see[10], and we have that

$$V_0(\lambda) = \frac{1}{n}\sum_{k=1}^{n} \frac{[z_k - X_k\beta_\lambda]^2}{[1 - a_{kk}(\lambda)]^2}\omega_k(\lambda) \tag{18}$$

where $\omega_k(\lambda) = \dfrac{[1 - a_{kk}(\lambda)]^2}{[1 - \frac{1}{n}\sum_{j=1}^{n} a_{jj}(\lambda)]^2}$, which is the expression for $V(\lambda)$ above.

Asymptotically it has been shown that the λ which minimizes the GCV function, $V(\lambda)$, is a good estimate of the λ which minimizes the 'true' predictive mean square error

$$R(\lambda) = \frac{1}{n}\sum_{i=1}^{n}[X_i\beta_\lambda - X_i\beta]^2 \tag{19}$$

where β is the true regression parameter in the model. A discussion of these and other related issues can be found in Craven and Wahba[6], Li[13], Lukas[15], Ragozin[19], Speckman[23] and Wahba[27, 28, 29]. Moreover it is known that the λ which minimizes $R(\lambda)$ frequently approximately minimizes other quadratic loss functions such as $\int [\theta_\lambda(s)-\theta(s)]^2 ds$ where θ_λ estimates the function θ, see Cox[5], Lukas[15] and Wahba[28].

3.2. The Non-Linear Case

When $\eta(\cdot, x_i)$ is non-linear, the Bayesian estimator is no longer a simple linear function of the data. However, by expanding $\eta(\beta, x_i)$ in a first order Taylor series expansion about β_λ, we can write β_λ in pseudo-linear form as

$$\beta_\lambda = [X'X + n\lambda\Sigma]^{-1}X'z^* \tag{20}$$

where $X_{ij} = \dfrac{\partial\eta(\beta, x_i)}{\partial\beta_j}|_{\beta=\beta_\lambda}$ and $z_i^* = X_i\beta_\lambda + z_i - \eta(\beta_\lambda, x_i)$. To extend the GCV function we now need to find an appropriate measure for degrees of freedom for error. A natural thing to do is look at the hat-matrix for the linearized estimator Eq.(20), *i.e.* $A(\lambda) = X(X'X + n\lambda\Sigma)^{-1}X'$, from which we construct an extended GCV function

$$V(\lambda) = \frac{\frac{1}{n}RSS(\lambda)}{[\frac{1}{n}tr(I - A(\lambda))]^2} \tag{21}$$

where $RSS(\lambda) = \sum_{i=1}^{n}[z_i - \eta(\beta_\lambda, x_i)]^2$. Again $V(\lambda)$ can be related to ordinary cross validation: With the same definition for $\beta_\lambda^{[k]}$, the ordinary cross validation function $V_0(\lambda)$, is given by:

$$V_0(\lambda) = \frac{1}{n}\sum_{k=1}^{n}[z_k - \eta(\beta_\lambda^{[k]}, x_k)]^2. \tag{22}$$

Letting δ_k be $\eta(\beta_\lambda^{[k]}, x_k) - z_k$ and $\beta_\lambda[z_k+\delta_k]$ be defined as in the linear case, one can prove, by an argument paralleling Lemma 3.1 of[6], that

$$\eta(\beta_\lambda^{[k]}, x_k) = \eta(\beta_\lambda[z_k+\delta_k], x_k). \tag{23}$$

Hence

$$V_0(\lambda) = \frac{1}{n}\sum_{k=1}^{n}\frac{[z_k - \eta(\beta_\lambda^{[k]}, x_k)]^2}{[1-a_{kk}^{\cdot}(\lambda)]^2} \tag{24}$$

where

$$a_{kk}^{\cdot}(\lambda) = \frac{\eta(\beta_\lambda[z_k+\delta_k], x_k) - \eta(\beta_\lambda[z_k], x_k)}{\delta_k}. \tag{25}$$

Approximating $a_{kk}^{\cdot}(\lambda)$ by the k'th diagonal element of the linearized hat-matrix, $a_{kk}(\lambda)$, we have

$$V_0(\lambda) \approx \frac{1}{n}\sum_{k=1}^{n}\frac{[z_k - \eta(\beta_\lambda, x_k)]^2}{[1 - a_{kk}(\lambda)]^2}, \tag{26}$$

and a relationship between $V(\lambda)$ and ordinary cross validation is obtained as before. By analogy with the linear case, one would suspect that the minimizer of $V(\lambda)$ ought to come close to minimizing the true predictive mean square error $R(\lambda) = \frac{1}{n}\sum_{i=1}^{n}[\eta(\beta_\lambda, x_i) - \eta(\beta, x_i)]^2$. In the next section we provide some numerical justification for this.

4. Remote Sensing Atmospheric Temperature Profiles

4.1. Description of the Problem

In order to illustrate the performance of our automatic Bayesian retrieval algorithm, we consider a remote sensing problem arising in satellite meteorology. Modern meteorological satellites, such as those in the TIROS-N series, have high resolution instruments on board,

which measure the intensity of up-welling radiation at selected channel frequencies. The basic features of these measurement systems are described in Smith *et al*[22]. The inversion problem is to retrieve an estimate of the atmospheric temperature profile, beneath the satellite, from the noisy upwelling radiance measurements. To proceed one needs to have some understanding of how atmospheric temperature profile information is manifested in the radiance data. For a non-scattering atmosphere in local thermodynamic equilibrium the radiative transfer equations, Liou[14], describe how the satellite up-welling radiance measurements relate to the underlying atmospheric temperature distribution T:

$$R_\nu(T) = B_\nu(T(x_s))\tau_\nu(x_s) - \int_{x_0}^{x_s} B_\nu[T(x)]\tau_\nu'(x)\,dx \tag{27}$$

where x is some monotone transformation of pressure p; x_s corresponding to the surface and x_0 corresponding to the top of the atmosphere. Meteorologists often work in *kappa* units, *i.e.* $x(p) = p^{5/8}$, since atmospheric variations are believed to be roughly constant in this scale. $\tau_\nu(x)$ is the transmittance of the atmosphere above x at wavenumber ν, and B_ν is Plank's function given by:

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

where

$$c_1 = 1.1906\times10^{-5} erg-cm^2-\sec^{-1}$$

and

$$c_2 = 1.43868\, cm-deg(K). \tag{28}$$

To an approximation one can use the radiative transfer equations to model satellite radiance data as:

$$y_i = R_{\nu_i}(T) + \epsilon_i \qquad i=1,2,\cdots,m \tag{29}$$

where y_i is the satellite radiance measurement at wavenumber ν_i in units of $mW/(m^2sr^1cm^{-1})$. The errors ϵ_i having mean zero and standard deviation $w_i^{-1/2}\sigma$. The relative weights, w_i, are

known but the scale factor σ is usually not known. In practice one typically has a good starting guess, T_o, for the temperature profile, either from the regional climatology or a meteorological forecast. Consequently the problem is to estimate an update or correction, θ, to this initial profile on the basis of given radiance data. In the notation of section 1, the objective is to estimate θ given measurements

$$y_i = N(\theta,\nu_i) + \epsilon_i \qquad i=1,2,\cdots,m \tag{30}$$

where $N(\theta,\nu_i) = R_\nu(T_o+\theta)$ and ϵ_i are mean zero random errors with standard deviation $w_i^{-1/2}\sigma$.

4.2. Monte Carlo Evaluation of the Retrieval Algorithm

We decided to study the behavior of the retrieval algorithm on the TIROS-N measurement system. The basic characteristics of the fifteen temperature sounding channels on this system are given in table I. These channels correspond to HIRS channels 1-7 and 13-17, and MSU channels 2-4 given in[22].

A more comprehensive description of these channels can be found in the Smith *et al* paper. To study the performance of the algorithm we conducted a small Monte Carlo experiment. Three real atmospheric temperature profiles, obtained directly from radiosones, and corresponding atmospheric transmission characteristics were kindly provided to us by Dr Thomas Koehler of the Meteorology Department at the University of Wisconsin - Madison. These profiles and the regional climatology are graphed in figure 1. For each profile, T, ten sets of simulated radiance measurements were generated according to

$$y_i = R_{\nu_i}(T) + \epsilon_i \qquad i=1,2,\cdots,15 \tag{31}$$

where ϵ_i is drawn from a normal distribution with mean zero and variance w_i^{-1}. Working from the simulated radiance data and using climatology as the initial guess, the retrieval algorithm was used to estimate the underlying temperature profile. The prior functional J which we used was $J(\theta) = \int_{x_0}^{x_s} [\theta''(x)]^2 dx$, reflecting the fact that apriori we thought the correction, resolvable

Table I Characteristics of the Temperature Sounding Channels on TIROS-N.

Channel Number	Wavenumber ν_i	Noise Level $w_i^{-1/2}$
1	668	2.9257
2	679	1.4018
3	691	1.5142
4	704	0.8697
5	716	1.1538
6	732	1.1443
7	748	1.6995
8	2190	0.01706
9	2213	0.01088
10	2240	0.01392
11	2276	0.01682
12	2361	0.02533
13	1.792	0.1595×10^{-4}
14	1.833	0.1391×10^{-4}
15	1.933	0.3092×10^{-4}

Figure 1 -ABOUT HERE-

by the instrument, should be smooth. Retrievals were done in a space of K=29 cubic B-splines which had knots of multiplicity three at the surface(p=1000 mb) and top(p=0 mb) of the atmosphere. The remaining 27 interior knots were roughly equispaced in kappa units. One wants to use a sufficient number K of B-splines so that no important resolution is lost due to approximation error. In general for a picture to be drawn on 8½×11 page, K=40 to 80 corresponds to the resolution of plotter pens and the human eye. However, the quality of the measurement system must also be taken into account. In the present application, due to the extreme ill-posedness of the problem, relatively little fine scale structure can be recovered. Consequently, it is believed that the choice of K=29 B-splines represented no tangible loss of accuracy over solving the problem with, say K=80. Our computations made extensive use of B-spline software originally written by De Boor[8].

The performance of the algorithm can be assessed in a variety of ways. First we look at the RMS errors in the estimated profile and compare these errors to the corresponding errors in the first guess. The RMS error for an estimated profile, $\hat{T}$, is defined to be

$$\text{RMS} = \sqrt{\frac{1}{15}\sum_{i=1}^{15} w_i [R_{\nu_i}(\hat{T}) - R_{\nu_i}(T)]^2} \tag{32}$$

where T is the true temperature profile; the RMS for the initial guess profile is similarly defined. Table II gives the median of the ten RMS values obtained on each of the three temperature profiles. One can see that the retrieval algorithm achieves a substantial reduction in this RMS error in all cases. The efficacy of the GCV procedure is measured by looking at the square of the ratio of the RMS error for the best possible choice of λ, to the RMS error corresponding to the value of λ minimizing the GCV *i.e.*

$$\text{efficacy} = [\min_{\lambda} \text{RMS}(\lambda)/\text{RMS}(\hat{\lambda})]^2 \tag{33}$$

Table II summarizes these efficacy numbers; for each profile the median efficacy is very close to 1.00 demonstrating that the GCV procedure does a good job of picking a value of λ which minimizes the RMS prediction error.

Table II Performance of the Retrieval Algorithm

Profile	RMS		Efficacy of GCV
	Initial Guess	Retrieved Estimate	
I	3.26	.65	.98
II	3.24	.58	.96
III	7.83	.57	.95

The bias and variability characteristics of the retrieved estimates are plotted in the figure 2.

Figure 2 -ABOUT HERE-

These bias and variability measures were constructed as follows: For a particular atmospheric level, x, corresponding to the ten sets of simulated radiance data, we have ten differences, e_i, between the estimated and true temperature profiles

$$e_i = \hat{T}_i(x) - T(x) \qquad i=1,2,....,10. \tag{34}$$

Since the distribution of these errors tends to be somewhat non-Gaussian, bias is defined to be the median, rather than the mean, of these differences and variability to be the median absolute deviation (MAD) of the differences divided by .6745. The factor .6745 is used so that if the e_i's were Gaussian then the variability measure would consistently estimate the standard deviation.

$$\begin{aligned} \text{bias} &= \text{med}\{\, e_i \quad , \; i=1,2, \cdots ,10 \,\} \\ \text{variability} &= \text{med}\{\, \frac{|e_i - \text{bias}|}{.6745} \, , \; i=1,2, \cdots ,10 \,\}. \end{aligned} \tag{35}$$

Figure 3 shows a retrieved profile and the corresponding true profile from a trial based on Profile III. The 5'th best case (based on RMS error) was chosen for display.

Figure 3 -ABOUT HERE-

Figures 2 and 3 indicate that the retrieval algorithm performs quite well. In all cases the bias is on the order of one degree while the variability is on the order of two to three degrees. These results are encouraging especially when one considers that we are using only fifteen noisy integral measurements. In fact if the radiative transfer equations are linearized about a standard atmosphere, approximate "transmission windows", W_i, for the fifteen channels can be found. These windows are plotted in figure 1 of Smith *et al*[22]. Channel measurements can roughly be represented as:

$$N(\theta, x_i) \approx \int W_i(x)\theta(x)\, dx. \tag{36}$$

Smith's plot shows that the transmission windows are broad and overlap, making the retrieval problem quite ill-posed. A reasonably good starting guess is required in this application but this

is typically available. The modest improvements over the starting guess which we obtain above are, economically, very important. Incorporating additional information such as the location of the tropopause or surface information ought to further improve the performance of the method. Even with the small number of measurements the GCV clearly emerges as a valuable technique for choosing the smoothing parameter.

References

1. R. Anderssen, F. de Hoog, and M. Lukas, *The Application and Numerical Solution of Integral Equations,* pp. 123-136, Sijthoff & Noordhoff, Germantown, 1980.

2. G. Backus and F. Gilbert, "Uniqueness in the inversion of inaccurate gross earth data," *Philos. Trans. Royal Soc. Ser. A.*, vol. 266, pp. 123-192, 1970.

3. C. T. H. Baker and G. F. Miller, *Treatment of Integral Equations by Numerical Methods,* Academic Press, London, 1982.

4. M. Chahine, "Generalization of the relaxation method for the inverse solution of non-linear and linear transfer equations," in *Inversion Methods in Atmospheric Remote Sensing,* ed. A. Deepak, Academic Press, New York, 1977.

5. D. D. Cox, "Approximation of the method of regularization estimators," Tech. Rep. No. 723, Statistics Dept., University of Wisconsin-Madison, 1983.

6. 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.*, vol. 31, pp. 377-403, 1979.

7. J. Crump and J. Seinfeld, "A new algorithm for the inversion of aerosol size distribution data," *Aerosol Science and Technology,* vol. 1, pp. 15-36, 1982.

8. C. DeBoor, *A Practical Guide to B-Splines,* Springer-Verlag, New York, 1978.

9. A. Deepak, *Inversion Methods in Atmospheric Remote Sensing,* Academic Press, New York, 1977.

10. G. Golub, M. Heath, and G. Wahba, "Generalized cross validation as a method for choosing a good ridge parameter," *Technometrics,* vol. 31, pp. 315-224, 1979.

11. W. C. Groetsch, "The Theory of Tikhonov Regularization for Fredholm equations of the First Kind," in *Research Notes in Mathematics 105,* Pitman, Boston, 1984.

12. M. Hutchinson and R. Bischof, "A new method for estimating the spacial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales," *Aust.*

Met. Mag., vol. 31, pp. 179-184, 1983.

13. Ker Chau Li, "From Stein's unbiased risk estimates to the method of generalized cross validation," Tech. Rep. No. 83-34, Statistics Dept., Purdue University, 1983.

14. K. Liou, *Introduction to Atmospheric Radiation,* Academic Press, London, 1979.

15. M. Lukas, "Regularization of linear operator equations.," Unpublished Ph.D. thesis, Australian National University, 1981.

16. P. Merz, "Determination of adsorption energy distribution by regularization and a characterization of certain adsorption isotherms," *J. Comp. Physics*, vol. 38, pp. 64-85, 1980.

17. D. Nychka, G. Wahba, S. Goldfarb, and T. Pugh, "Cross-validated spline methods for the estimation of three dimensional tumor size distributions from observations on two dimensional cross sections," Tech. Rep. No. 711, Statistics Dept., University of Wisconsin-Madison, 1983. To appear, *J. Am. Stat. Soc.*.

18. F. O'Sullivan, "On the analysis of some penalised likelihood estimation schemes," Tech. Rep. No. 726, Statistics Dept., University of Wisconsin-Madison, 1983.

19. D. Ragozin, "Error bounds for derivative estimates based on spline smoothing of exact or noisy data," Tech. Rep. No. GN-50, Statistics Dept., University of Washington-Seattle, 1981.

20. L. B. Rall, *Computational Solution of Nonlinear Operator Equations,* John Wiley & Sons, 1969.

21. W. Smith, "The retrieval of atmospheric profiles from VAS geostationary radiance observations," *J. Atmospheric Sciences*, vol. 40, pp. 2025-2035, 1983.

22. W. L. Smith, H. M. Woolf, C. M. Hayden, D. Q. Wark, and L. M. McMillin, "The TIROS-N operational vertical sounder," *Bull. Amer. Meteor. Soc.*, vol. 50, no. 10, pp. 1177-1187, October 1979.

23. P. Speckman, "Efficient nonparametric regression with cross-validated smoothing splines," Tech. Rep. No. 45, Statistics Dept., University of Missouri-Columbia, 1982.

24. J. Surmont and Y. M. Chen, "Numerical solutions of a nonlinear radiative transfer equation with inadequate data," *J. Comp. Physics*, vol. 13, pp. 288-302, 1973.

25. A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems,* Wiley, New York, 1977.

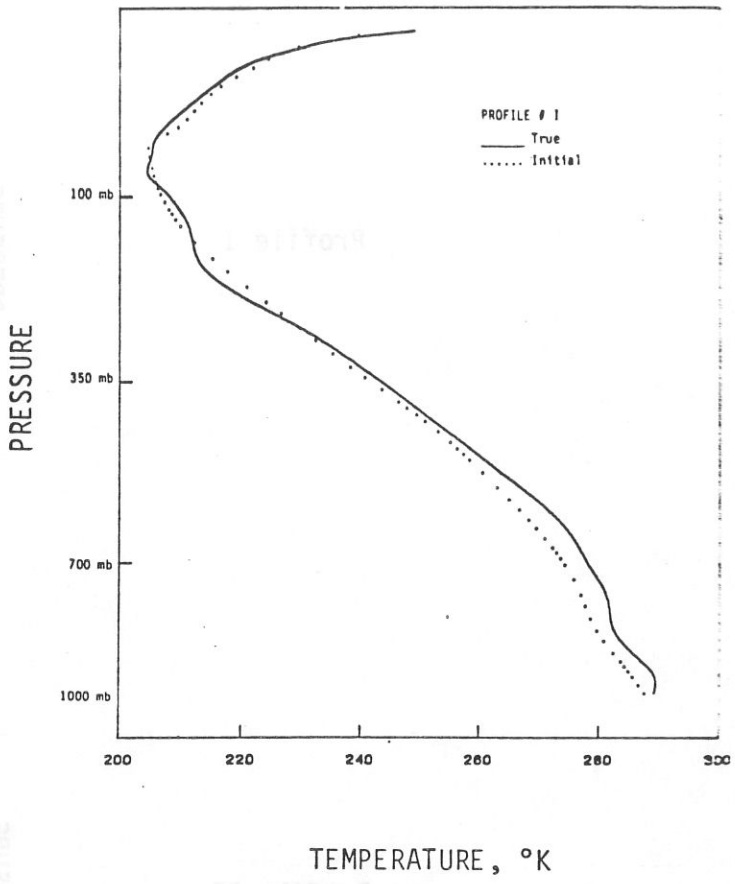
26. S. Twomey, *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements,* Elsevier, New York, 1977.

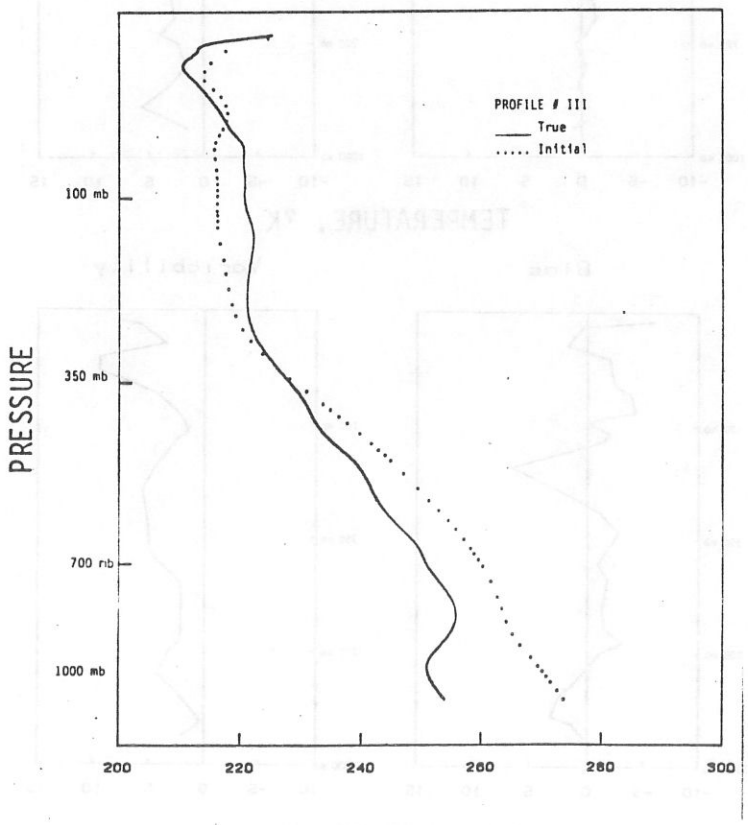
27. G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy," *SIAM J. Numer. Anal.*, vol. 14, pp. 651-667, 1977.

28. G. Wahba, "Smoothing and ill-posed problems," in *Solution Methods for Integral Equations with Applications*, ed. M. Goldberg, pp. 183-194, Plenum Press, 1979.

29. G. Wahba, "Constrained regularization for ill-posed linear operator equations, with applications in meteorology and medicine," in *Statistical Decision Theory and Related Topics III*, ed. J. O. Berger and S. S. Gupta, pp. 383-418, Academic Press, 1982.

Figure 1: True and Initial (regional climatology) temperature profiles I, II, and III.





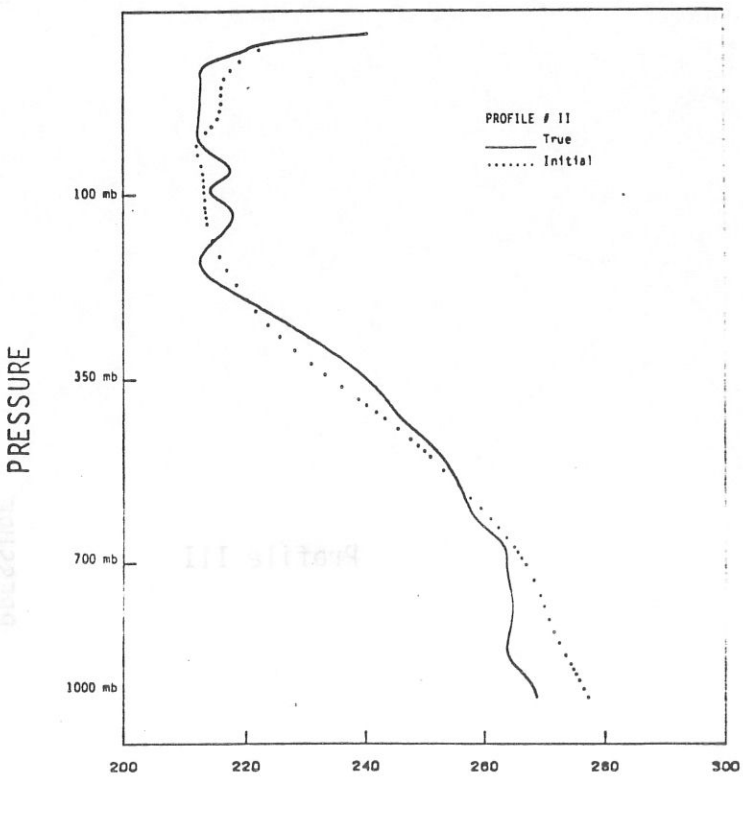
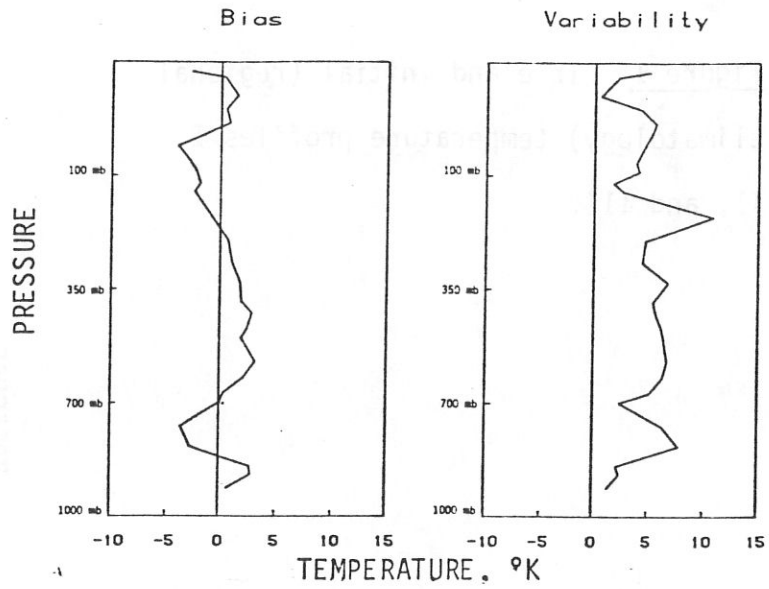
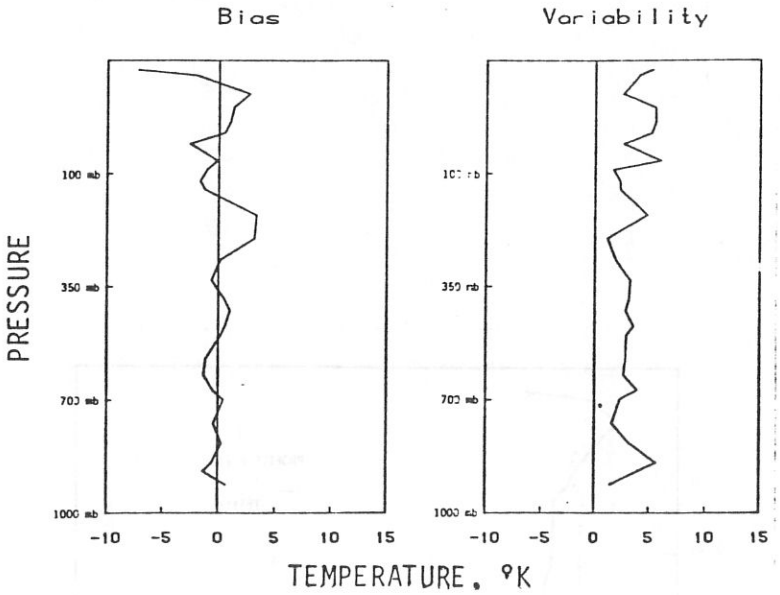


Figure 2: Bias and Variability Characteristics of Retrievals

Profile I



Profile II



Profile III

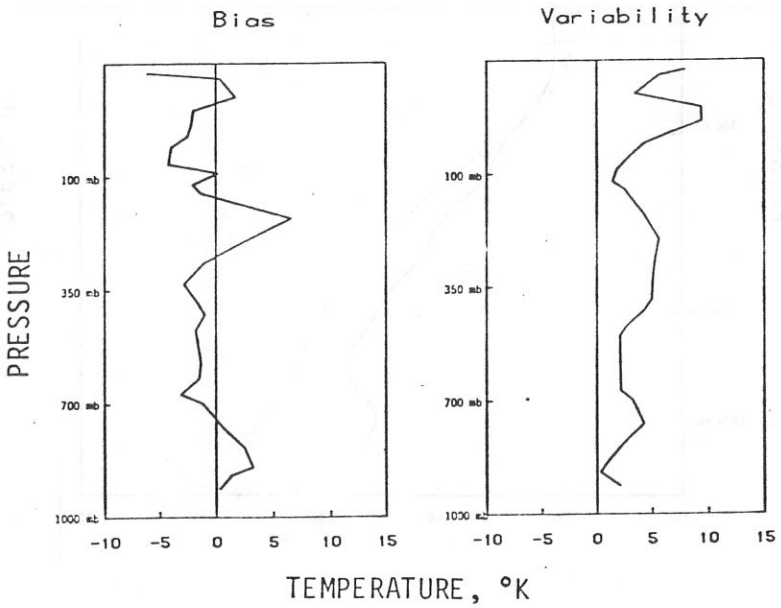


Figure 3: Fifth best retrievals, with respect to an RMS performance ordering.