DEPARTMENT OF STATISTICS

University of Wisconsin
1210 W. Dayton St.
Madison, WI 53706

TECHNICAL REPORT NO. 692

August 1982

VARIATIONAL METHODS IN SIMULTANEOUS OPTIMUM INTERPOLATION AND INITIALIZATION

Grace Wahba

Statistics Department
University of Wisconsin-Madison

This research supported by NASA Grant NAG5-128.

VARIATIONAL METHODS IN SIMULTANEOUS OPTIMUM INTERPOLATION AND INITIALIZATION

Grace Wahba
Statistics Department
University of Wisconsin-Madison

ABSTRACT

This report was prepared for the Proceedings of the Fourteenth Stanstead Seminar, sponsored by the Department of Meteorology, McGill University, and the National Center For Atmospheric Research, and was written with the Stanstead audience in mind. The theme of the Seminar was "The interaction between objective analysis and initialization". We first review a duality between optimum interpolation and variational objective analysis. We use this duality to set up a variational approach to objective analysis which uses prior information concerning the atmospheric spectral energy distribution, in the variational problem. In the wind analysis example we study, the wind field is partitioned into divergent and nondivergent parts, and a control parameter governing the relative energy in the two parts is estimated from the observational data being analyzed by generalized cross validation, along with a bandwidth parameter. We then propose a variational approach to combining objective analysis and initialization in a single step. In a simple example of this approach, data, forecast, and prior information concerning atmospheric energy distribution is combined into a single variational problem. This problem has (at least) one bandwidth parameter, one partitioning parameter governing the relative energy in fast and slow modes, and one parameter governing the relative weight to be given to observational and forecast data. It may be possible to estimate good values of all three parameters from the observational and forecast data.

1. INTRODUCTION

In this work, we exploit a duality between optimum interpolation and variational objective analysis to set up a certain variational approach to objective analysis, as well as to propose an approach to combining objective analysis and variational initialization into a single step.

In Section 2 we describe the duality. In Section 3 we describe the variational approach to objective analysis in the context of the estimation of vorticity and divergence from observed wind vectors. The duality of Section 2 is used to suggest how data concerning the energy spectral distribution in the atmosphere, as studied by, e.g., Baer (1974, 1981), Kasahara (1976), Kasahara and Puri (1981), Stanford (1979), may be used to help choose the form of the variational problem to be solved. The variational problem we solve to estimate vorticity and divergence has one "bandwidth" parameter (related to the half power point of the equivalent low pass filter) and one "partitioning" parameter, representing the relative allocation of energy to the divergent and non-divergent part of the wind. Both can be estimated by generalized cross validation (GCV).

In Section 4 we begin a synthesis of several ideas - i) the duality of Section 2, ii) the idea of partitioning the "signal" into divergent and non-divergent parts (which generalizes to the idea of partitioning the signal into slow and fast modes) and iii) the modified Kalman filter ideas as proposed by Ghil and coworkers (1981). The result is an argument that (in principle) one can set up the problem of estimating initial conditions by combining i) the forecast, ii) the observational data, iii) possibly certain physical constraints and iv) (partial) prior information concerning atmospheric spectral energy distribution, into a single variational problem. This result is in some sense a converse of Phillips (1982b), who argues that normal mode initialization can be done as part of optimum interpolation. The resulting variational problem as we propose it will have (at least) one bandwidth parameter, one balance parameter controlling the relative weight to be given to forecast data and observational data, and one partitioning parameter, governing the relative energy in the "signal" assigned to fast and slow modes. We conjecture that these three (control) parameters can be estimated dynamically from the data by GCV. They can also be chosen by trial and error. Objective analysis and (linear) normal mode initialization are thereby combined in one step. The estimation of one (or more) balance parameter(s) from the data may possibly avoid the pitfalls due to misspecification of the Kalman filter statistics as noted by Phillips (1982a).

Some of this work is joint with D. R. Johnson.

2. ON A DUALITY BETWEEN OPTIMUM INTERPOLATION AND VARIATIONAL OBJECTIVE ANALYSIS

We will describe this duality for the analysis of a univariate variable on the sphere (say 500 mb height) although the result is completely general. Let P denote a point on the sphere and let h(P) be (say), the 500 mb height minus the global average 500 mb height at P. Let the observations $y_1, \ldots y_n$ be modelled as

$$y_i = h(P_i) + \varepsilon_i \quad , \tag{2.1}$$

where the ε_i are supposed to be zero mean independent measurement errors with common variance $\sigma^2 = E\,\varepsilon_i^2$.

We suppose E h(P) = 0 and h has a prior covariance b R(P,Q),

$$E\, h(P)h(Q) = b\, R(P,Q).$$

Then using standard results in multivariate analysis

$$E\big(h(P)|y_1 \ldots y_n\big) = \big(R(P,P_1)\ldots R(P,P_n)\big)(R_n + n\lambda I)^{-1} \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \tag{2.2}$$

$$= h_\lambda(P), \text{ say}$$

where $\lambda = \sigma^2/nb$ and R_n is the n x n matrix with ij th entry $R(P_i,P_j)$. Considering the expression on the right of (2.2) as a function of P, which we denote by $h_\lambda(P)$, it is easy to see that if $\lambda = 0$, then $h_0(P)$ interpolates to the data exactly, $h_0(P_i) = y_i$, $i = 1,2,\ldots n$, whereas for $\lambda > 0$, $h_\lambda(P)$ smooths the data, and λ controls the amount of smoothing -λ is the "bandwidth" parameter, $h_\lambda(P)$ evaluated at grid points is the "optimum interpolant" of Gandin, given that all the available data points are used simultaneously.

Duality Theorem: Kimeldorf and Wahba (1970, 1971) Wahba (1978).

For every covariance R(P,Q) satisfying $\int\int R^2(P,Q)dPdQ<\infty$, there is a variational problem for which $h_\lambda(P)$ is the solution. It is: find h in H_R (a certain reproducing kernel Hilbert space) to minimize

$$\frac{1}{n}\sum_{i=1}^{n} \big(y_i - h(P_i)\big)^2 + \lambda\, J(h). \tag{2.3}$$

where J(h) is the square norm of h in H_R.

We will now describe J(h) in a manner which we hope will make its meteorological usefulness clear. R(P,Q) being a covariance, is a symmetric non-negative definite function and (given that it satisfies the hypothesis) has a so called Mercer-Hilbert-Schmidt expansion (Riesz -Sz.-Nagy (1955)) in its eigenfunctions and eigenvalues

$$R(P,Q) = \sum_{\ell,s} \lambda_{\ell s} Y_{\ell s}(P) Y_{\ell s}(Q) , \text{ where } \int R(P,Q) Y_{\ell s}(Q)dQ = \lambda_{\ell s} Y_{\ell s} (P). \quad (2.4)$$

(This expansion is a generalization of the factorization of a covariance matrix Σ as $\Sigma = \Gamma D \Gamma'$ where $\Gamma\Gamma' = I$ and D is diagonal). We have deliberately used a notation to suggest that the eigenfunctions of R are spherical harmonics, but that is by no means necessary. In any case

$$J(h) = \sum_{\ell s} \frac{h_{\ell s}^2}{\lambda_{\ell s}} , \quad \text{where} \quad h_{\ell s} = \int h(P) Y_{\ell s}(P)dP. \quad (2.5)$$

(If h were a vector and R a matrix, then we would have $J(h) = h'R^{-1}h$). As an example, if the $Y_{\ell s}$ are the spherical harmonics and $\lambda_{\ell s} = [\ell(\ell+1)]^{-2m}$ then by using the fact that the spherical harmonics are the eigenfunctions of the Laplacian, $\Delta Y_{\ell s} = -\ell(\ell+1) Y_{\ell s}$, it is not hard to show that $J(h) = \iint(\Delta^m h)^2 dP$.

More generally, if $\lambda_{\ell s} = |\sum_{\nu=0}^{m} \alpha_\nu[(\ell)(\ell+1)]^\nu|^{-2}$, then

$$J(h) = \int| \sum_{\nu=0}^{m} \alpha_\nu (-\Delta)^\nu h|^2 dP. \quad (2.6)$$

More details may be found in Wahba (1981a, 1981b), the use of Hough functions instead of spherical harmonics is briefly described in Wahba (1981b). If

$$E\ h(P)h(Q) = b\ R(P,Q) \quad , \quad (2.7)$$

then h has a Karhunen-Loeve expansion

$$h(P) = b \sum h_{\ell s} Y_{\ell s}(P) \quad \text{where} \quad bh_{\ell s} = \int h(P) Y_{\ell s} (P)\ dP , \quad (2.8)$$

and the $h_{\ell s}$ are independent zero mean random variables with $E\ h_{\ell s} h_{\ell' s'} = b\lambda_{\ell s}$, $\ell,s = \ell',s'$, $= 0$ otherwise. The $\{\lambda_{\ell s}\}$, that is, the relative energy distribution by wave numbers may be estimated from historical data and/or obtained (roughly) from theory, see Baer (1974, 1981), Kasahara (1976), Kasahara and Puri (1981), Stanford (1979).

3. ESTIMATION OF DIVERGENCE AND VORTICITY FROM THE OBSERVED WIND FIELD USING VECTOR SPLINES ON THE SPHERE. PARTITIONING OF THE DATA INTO DIVERGENT AND NON-DIVERGENT PARTS

Given observed wind data (u_i,v_i) at point P_i, $i = 1,2,. . . n$, we estimate the vorticity and divergence as follows. The stream function and velocity potential are expanded in spherical harmonics

$$\Psi (P) = \sum_{\ell=1}^{L} \sum_{s=-\ell}^{\ell} a_{\ell s} Y_{\ell s}(P) , \quad \Phi (P) = \sum_{\ell=1}^{L} \sum_{s=-\ell}^{\ell} b_{\ell s} Y_{\ell s}(P) . \quad (3.1)$$

Then (for given δ,λ), we find $\{a_{\ell s}, b_{\ell s}\}$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}\left(-\frac{1}{a}\frac{\partial\Psi}{\partial\phi}(P_i)+\frac{1}{a\cos\phi_i}\frac{\partial\Phi}{\partial\lambda}(P_i)-u_i\right)^2 \tag{3.2}$$

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

$$+\lambda\left[J_1(\Psi)+\frac{1}{\delta}J_2(\Phi)\right]$$

where

$$J_1(\Psi)=\sum_{\ell=1}^{L} a_{\ell s}{}^2/\lambda_{\ell s}{}^{(1)},\quad J_2(\Phi)=\sum_{\ell=1}^{L} b_{\ell s}{}^2/\lambda_{\ell s}{}^{(2)}. \tag{3.3}$$

Given estimates of the $a_{\ell s}$ and $b_{\ell s}$, estimates of the wind, vorticity and divergence are obtained analytically for any desired P. The resulting estimated wind field is called a vector spline on the sphere. λ is the bandwidth parameter, it controls the partitioning of the data vector $(u_1, \ldots u_n, v_1, \ldots v_n)'$ into a part due to "signal" and a part due to noise. δ can be viewed as a signal partitioning parmeter, it controls the partitioning of the "signal" part of the data into a divergent part and a non-divergent part.

A Monte Carlo study was carried out to test the effectiveness of the method and to determine whether good estimates of λ and δ could be obtained from the data by GCV. The results are reported in Wahba (1982), and in preparation. The weights $\lambda_{\ell s}{}^{(1)}$ and $\lambda_{\ell s}{}^{(2)}$ were adapted from the data collected by Stanford (1979). Realistically scaled model 500 mb wind fields were generated via a Karhunen-Loeve expansion in stream function and velocity potential, the resulting "true" wind vectors at 114 North American weather stations computed and a random 2.5 m/sec rms measurement error added to each computed wind component. Good recovery of winds, vorticity and divergence in an area covered by the data grid and extending a small amount past it was obtained using the estimated λ and δ. (The individual $a_{\ell s}$, $b_{\ell s}$ are not recovered from only North American data). The results are sensitive to both λ and δ. It can be seen from the experiments that a poor value of δ causes obvious edge effects and that fixing δ at values which oversuppress the divergent part of the wind tend to cause increased errors in the estimation of the non-divergent part.

4. ON A SINGLE VARIATIONAL PROBLEM FOR MERGING DATA, FORECAST AND PRIOR KNOWLEDGE OF ATMOSPHERIC SPECTRAL ENERGY DISTRIBUTION

We claim that a set of "moderately" reasonable assumptions concerning the forecast errors and measurement errors, along with prior knowledge concerning the spectral energy distribution in the atmosphere, leads to an "optimal" initial state obtained as the solution of a single (large) variational problem. This variational problem will have the same number of "unknowns" as the degrees of freedom in the forecast model, thus raising questions of numerical feasibility. We brush aside computational questions for the present, adopting the point of view that it is worthwhile to examine an "optimal" variational problem, and then ask, how close can one come to computing a reasonable approximation to it. (Some numerical shortcuts appear in Bates and Wahba (1982)). One of the advantages of examining the variational form is that it is fairly evident how to add side constraints based on the physics.

A (simplified) example goes as follows. Consider a spectral model where the state of the atmosphere is expressed in terms of Hough functions. Using notation similar to Tribbia (1982), and considering a single level, let

$$\begin{pmatrix} U \\ V \\ \phi \end{pmatrix} = \sum_{j=1}^{N_{max}} X_j \underset{\sim}{H}_j^R + \sum_{k=1}^{M_{max}} Y_k \underset{\sim}{H}_k^G \tag{4.1}$$

(U,V,ϕ) is the "true" wind and geopotential field, and $\underset{\sim}{H}_j^R$ and $\underset{\sim}{H}_k^G$ are rotational (slow) and gravitational (fast) Hough functions. We suppose that the true "state" of the atmosphere is adequately described by the $N_{max} + M_{max}$ vector $\theta = (\underset{\sim}{X}:\underset{\sim}{Y}) = (X_1, \ldots X_{N_{max}} : Y_1, \ldots Y_{M_{max}})$ and an analysis consists of obtaining an updated estimate $\hat{\theta}$ of θ, given a forecast

$$\theta^F = (\underset{\sim}{X}^F:\underset{\sim}{Y}^F) = (X_1^F, \ldots, X_{N_{max}}:Y_1^F, \ldots, Y_{M_{max}}^F), \tag{4.2}$$

given data (u_i, v_i, ϕ_i) representing observations on the wind and geopotential height at point P_i, and given the "prior knowledge" concerning atmospheric energy distribution obtained, e.g. from data like that collected by Kasahara and Puri (1981). This prior knowledge is of the form

$$EX_j^2 = b_1 \lambda_j^R = EY_k^2 = b_2 \lambda_k^G . \tag{4.3}$$

where we assume that the relative energies λ_j^R and λ_k^G are known but that b_1 and b_2 are not. We will assume that all cross covariances $E X_i X_j$, $E X_i Y_j$ and $E Y_i Y_j$ can be approximated by 0. Suppose further that

$$E(X_j-X_j^F)(X_k-X_k^F) = \omega_F \sigma_{jk}^{XX} , \quad E(Y_j-Y_j^F)(Y_k-Y_k^F) = \omega_F \sigma_{jk}^{YY} \tag{4.4}$$

$$E(X_j-X_j^F)(Y_k-Y_k^F) = \omega_F \sigma_{jk}^{XY}$$

where the σ_{jk}'s are known, and let Σ be the $(N_{max} + M_{max}) \times (N_{max} + M_{max})$ covariance matrix with entries σ_{jk}^{XX}, σ_{jk}^{XY}, σ_{jk}^{YY}. Let the measurement errors of (u_i, v_i, ϕ_i) be independent with variances $\omega_o\sigma_u^2$, $\omega_o\sigma_v^2$, $\omega_o\sigma_\phi^2$. Suppose all random variables are normally distributed. Then

<u>Theorem:</u> (G. Wahba, in preparation) The Bayes estimate $\hat{\theta} = (\hat{\underset{\sim}{X}} : \hat{\underset{\sim}{Y}})$ of $\theta = (\underset{\sim}{X} : \underset{\sim}{Y})$ is the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}\left\| \begin{matrix} u_i \\ v_i \\ \phi_i \end{matrix} - \sum_{j=1}^{N_{max}} X_j \underset{\sim}{H}_j^R(P_i) - \sum_{k=1}^{M_{max}} Y_k \underset{\sim}{H}_k^G(P_i) \right\|_\sigma^2 \tag{4.5}$$

$$+ \omega\{(\underset{\sim}{X}-\underset{\sim}{X}^F : \underset{\sim}{Y}-\underset{\sim}{Y}^F)\,\Sigma^{-1}\,(\underset{\sim}{X}-\underset{\sim}{X}^F : \underset{\sim}{Y}-\underset{\sim}{Y}^F)'\}$$

$$+ \lambda\left\{\sum_{j=1}^{N_{max}} \frac{X_j^2}{\lambda_j^R} + \delta \sum_{k=1}^{M_{max}} \frac{Y_k^2}{\lambda_k^G}\right\}$$

where $\left\|\begin{matrix} u \\ v \\ \phi \end{matrix}\right\|_\sigma^2 = \frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} + \frac{\phi^2}{\sigma_\phi^2}$, and $\omega = \omega_o/\omega_F n$, $\lambda = \omega_o/b_1 n$, $\delta = b_1/b_2$.

Since the expression to be minimized is a quadratic form in the components of $(\underset{\sim}{X}:\underset{\sim}{Y})$, (for given ω, λ, δ), the minimizer can be readily expressed as the solution to a linear system. It is conjectured that the bandwidth parameter λ, the partitioning parameter δ and the "error balancing" parameter ω, can all be estimated (simultaneously!) from the data by GCV. Note that a large ω suppresses forecast relative to observational data and a large δ suppresses fast modes relative to slow modes. The results reported in Section 3 suggest that at least one bandwidth and one (appropriately chosen) partitioning parameter can be chosen by GCV. Part of the art of this approach is to choose these "tuning" parameters so that they are (a) the important ones and (b) their determination is relatively "well posed." Sensitivity of optimum interpolation to bandwidth parameters has recently been discussed by Hollingsworth (1982), Lorenc (1981) and others.

To be practical it may be necessary to assume an oversimplified structure for Σ and/or to break up the variational problem into a series of sub-problems. The usual Kalman filtering theory would give $\hat{\theta}$ as the minimizer of the above expression, with $\lambda = 0$ and with $(\omega_F\Sigma)$, the forecast error covariance, given by a recursion formula in time. Our description above, implicitly assumes the existence of a limiting ($\omega_F\Sigma$). The forecast of Ghil et al. would, roughly speaking reduce in this context , to setting $\lambda = 0$ and constraining $\hat{\underset{\sim}{Y}}$ to be 0. The above theorem generalizes to the simultaneous analysis of all levels. Then satellite radiance data (which "cuts across" all levels) can be included as another term in the variational formulation. Physical constraints can, in principle be included as side conditions. It also appears that non-linear balancing constraints related to those discussed in, e.g. Tribbia (1982), whose purpose is to minimize the propagation of unwanted gravity waves in the non-linear forecast equations, may also be incorporated as either weak or strong constraints.

REFERENCES

Baer, F., (1974) Hemispheric spectral statistics of available potential energy, J. Atmos. Sci. 31, 4, 932-941.

Baer, F., (1981) Three dimensional scaling and structure of atmospheric energetics, J. Atmos. Sci. 38, 1, 52-68.

Bates, D., and G. Wahba, (1982) Computational methods for generalized cross validation with large data sets. In preparation.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, (1981) Applications of estimation theory to numerical weather prediction, Dynamic Meteorology, L. Bengtsson, M. Ghil, E. Kallen, eds., Applied Mathematical Sciences 36, Springer Verlag, N. Y.

Hollingsworth, A., (1982) Operational data assimilation at ECMWF, these proceedings.

Kasahara, A., (1976) Normal modes of ultra-long waves in the atmosphere, Mon. Wea. Rev. 106, 669-690, Fig. 14.

Kasahara, A., and K. Puri, (1981) Spectral representation of three dimensional global data by expansion in normal mode functions, Mon. Wea. Rev. 109, 37-51.

Kimeldorf, G., and G. Wahba, (1970) A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, Ann. Math. Statist. 41, 495-502.

Kimeldorf, G., and G. Wahba, (1971) Some results on Tchebycheffian spline functions, J. Math. Anal. and Applic. 33, 1.

Lorenc, A. C., (1981) A global three dimensional multivariate statistical interpolation scheme, Mon. Wea. Rev. 109, 4, 701-721.

Phillips, N. A., (1981a) Variational analysis and the slow manifold, Mon. Wea. Rev. 109, 12, 2415-2426.

Phillips, N. A., (1982a) A very simple application of Kalman filtering to meteorological data assimilation, Office note 258, NMC.

Phillips, N. A., (1982b) On the completeness of multivariate optimum interpolation for large scale meteorological analysis, Manuscript.

Riesz, F., and Sz.-Nagy, B., (1955) Functional Analysis, Ungar, N.Y., 242-246.

Stanford, J., (1979) Latitudinal-wavenumber power spectra of stratospheric temperature fluctuations, J. Atmos. Sci., 36,5, 921-931.

Tribbia, J. J., (1982) On variational normal mode initialization, Mon. Wea. Rev., 110, 6, 455-470.

Wahba, G., (1978) Improper priors, spline smoothing, and the problem of guarding against model errors in regression, J. Roy. Stat. Soc. B, 40, 364-372.

Wahba, G., (1981a). Spline interpolation and smoothing on the sphere, SIAM J. Scientific and Statistical Computing, 2, 1. Also Univ. of Wisconsin-Madison, Dept. of Statistics TR#584.

Wahba, G., (1981b). Some new techniques for variational objective analysis on the sphere using splines, Hough functions, and sample spectral data. Preprints of the Seventh Conference on Probability and Statistics in the Atmospheric Sciences, American Meteorological Society.

Wahba, G., (1982). Vector splines on the sphere, with application to the estimation of vorticity and divergence from discrete, noisy data, to appear in "Multivariate Approximation Theory," Vol. 2, W. Schempp and K. Zeller, eds., Birkhauser Verlag.