TECHNICAL REPORT NO. 818

January, 1988

On the dynamic estimation of relative weights for observation and forecast in numerical weather prediction.

by

Grace Wahba


RSRM '87 Advances in Remote Sensing Retrieval Methods 347-358
Deepak, Fleming & Theon, Eds.

This research supported in part by NASA under Contract NAG5-315 and NSF under Grant ATM-840373.
ON THE DYNAMIC ESTIMATION OF RELATIVE WEIGHTS
FOR OBSERVATIONS AND FORECAST IN NUMERICAL WEATHER PREDICTION

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

ABSTRACT

We look at the problem of merging direct and remotely sensed (indirect) data with forecast data to get an estimate of the present state of the atmosphere, for the purpose of numerical weather prediction. To carry out this merging optimally, it is necessary to provide an estimate of the relative weights to be given to the observations and forecast. It is possible to do this dynamically from the information to be merged, if the correlation structure of the errors from the various sources is sufficiently different. We describe some new statistical approaches to doing this and quantify conditions under which such estimates are likely to be good.

1. INTRODUCTION

We have been studying various aspects of the problem of simultaneously combining information from various sources, for the purpose of obtaining initial conditions of the atmosphere for numerical weather prediction. By information, we mean data from diverse instruments, information from a forecast, prior information concerning the atmosphere, and physical constraints. See Wahba(1981,1982a, 1985a). Various parts of this what might be called "multispectral" point of view, whereby data from different sources is combined, and several meteorological parameters retrieved simultaneously, is a major theme in several papers presented at this conference, in particular, see Isaacs et al. (1988), Smith et al.(1988), Westwater et al. (1988) and Rodgers (1988). See also Lorenc (1986).

In this paper we will first briefly review the variational prescription for combining data from different sources and prior information, and its relation to Gandin (Bayes) estimation. In this prescription a number of "tuning" parameters, which we will divide into two classes. The first class will be called weighting parameters, and the second smoothing (also known as bandwidth) parameters. The weighting parameters are those which govern the relative weights to be given to various types of observational, forecast, and physical information, while the smoothing parameters control the relative amount of "information" which is to be assigned to "signal" and to "noise". All practical forecast models have many such tuning parameters, and in practice, they are chosen by trial and error, by the use of externally measured data on various sources of error and strength of signal, and in very simple cases, by Kalman filtering, which propagates estimates of covariances forward in the model.
Under certain circumstances certain smoothing parameters can be estimated well from
the data at hand (i.e. the data to be analyzed), by cross validation and generalized cross
validation (GCV) methods. These methods sit somewhere between "static" estimation
methods, with their long response times, and the Kalman filter methods, with their relatively
large computational burden, and the possibility that the assumptions of the Kalman filter
theory concerning error structure might not be satisfied for model errors in some cir-
cumstances.

It is the purpose of this paper to initiate the development of a complimentary theory for
the dynamic estimation of certain weighting parameters, which govern the relative weight to
be given to different types of observations and forecast. In practice, the estimates resulting
from this theory have the potential for being used to decide which of two strongly conflicting
sources of information (for example, forecast and satellite radiance data) should be given pri-
mary credence. This theory must of necessity have certain subtleties, since, if two instru-
ments (or an instrument and forecast) measure the same quantity, each with a constant bias,
the relative sizes of the two biases cannot be discerned from the observations (or observa-
tions and forecast). In order to have hope of carrying out this program, we shall see that the
spatial (or temporal) error correlation structure from the two different sources has to be
sufficiently different. Part of the goal of this theory is to quantify "sufficiently different" in a
useful way. We believe that this circumstance of "sufficiently different" occurs in a number
of meteorologically important circumstances.

After reviewing the general variational problem, we present the estimates, using 500mb
heights from raobs and forecast as a concrete example. Then we give a theorem which
quantifies "sufficiently different" and, finally, we describe the general case of both direct and
indirect measurements of the same meteorological quantities.

2. VARIATIONAL AND GANDIN OBJECTIVE ANALYSIS

We suppose that \( s_1, \ldots, s_p \) are the "state variables" in a global scale numerical weather
prediction model, either a grid point or a spectral model. We will assume that IF the best
possible values of the state variables were chosen, then the difference between the model
atmosphere and the true atmosphere can be treated as "noise". The number \( p \) of state vari-
ables may be very large.

We will consider two vectors of information (think of these as direct or indirect observa-
tional data, or forecast), that are to be combined to obtain an improved estimate of the state
variables.

Let \( y^{(1)} \) and \( y^{(2)} \) be vectors from set 1 and set 2 respectively and suppose
\[
y^{(i)} = F^{(i)}(s) + \epsilon^{(i)}, \quad i = 1, 2
\]
where \( y^{(i)} = (y_{1}^{(i)}, \ldots, y_{n_i}^{(i)}) \), \( s = (s_1, \ldots, s_p) \), \( F^{(i)}(s) = (F_{1}^{(i)}(s), \ldots, F_{n_i}^{(i)}(s)) \), \( i = 1, 2 \) and we
assume that the discrepancies \( \epsilon^{(i)} \) between \( y^{(i)} \) and \( F^{(i)} \) can be modelled as zero mean ran-
dom vectors with \( E\epsilon^{(i)}\epsilon^{(i)^T} = \sigma_i^2 Q_i, i = 1, 2 \), and \( E\epsilon^{(1)}\epsilon^{(2)} = 0 \), that is, errors in the two data
sources are uncorrelated.

\( F^{(i)} \) may be a matrix (including the identity matrix), or a nonlinear operator which
models the forward problem arising when a radiometer observes radiant energy remotely, or
it may model the relationship between directly measured quantities such as the horizontal wind field, and state variables in the forecast model such as the coefficients in a spherical harmonic expansion of stream function and velocity potential.

We will suppose that a reasonable model for the state variables (the mean having been subtracted off), is

\[ E_s = 0, \quad E_{ss'} = b\Sigma \]

See Wahba(1982b), for further references and a discussion as to how \( \Sigma \) may be obtained from historical data.

This approach results in a mandate to find \( s \) as the minimizer of the variational problem:

\[
(y^{(1)}(s) - F^{(1)}(s))Q_1^{-1}(y^{(1)} - F^{(1)}(s)) + \frac{1}{r}(y^{(2)} - F^{(2)}(s))Q_2^{-1}(y^{(2)} - F^{(2)}(s)) + \lambda s'\Sigma^{-1}s,
\]

where \( r = \frac{\sigma_2^2}{\sigma_1^2} \) and \( \lambda = \frac{\sigma_1^2}{b} \). See Wahba (1981, 1982a, 1985a), O'Sullivan and Wahba(1985).

We note that physical constraints on the state vector \( s \) or functions of \( s \) can be inserted by solving the variational problem above subject to these constraints, see Villalobos and Wahba(1982), Svensson(1985). Svensson put constraints on the dry adiabatic lapse rate when solving a variational problem of this form with satellite radiance data to estimate vertical temperature profiles.

In the linear Normally distributed case the estimate of \( s \) which minimizes the above variational problem is the Gandin(Bayes) estimate of \( s \), given the prior covariance of \( s \) and the covariance matrices of the errors. See, e.g. Kimeldorf and Wahba(1970). Kalman filter theory would give \( s \) as the minimizer of the variational problem with the term preceeded by \( \lambda \) absent. Here the inclusion of this term enters apriori (smoothness) information that may have been lost via model error.

If \( r, Q_1 \) and \( Q_2 \) are known, then \( \lambda \) and some parameters inside \( \Sigma \) may be estimated by the GCV See, for example, Wahba and Wendelberger(1980), Wahba(1982b), O'Sullivan and Wahba(1985), Merz(1980).

3. DYNAMIC ESTIMATION OF WEIGHTING PARAMETERS

3.1 A SIMPLE EXAMPLE - 500 mb RAOBS AND FORECAST

We first illustrate the method and results by letting \( y^{(1)} \) be observed 500mb heights and \( y^{(2)} \) be forecast 500mb heights. It is reasonable to take \( Q_1 \) to be \( I \) for the observations since these measurements can be assumed to be independent, with about the same variance from station to station. Hollingsworth and Lonnberg(1986) and Lonnberg and Hollingsworth(1986) (L.H), have recently obtained estimates of \( r \) and \( Q_f=Q_2 \) from three months data from the European Center forecast model. We will use their example and results as an illustration, before going on to the general case.
Figure 1. (From LH). The correlation of 500mb height forecast errors as a function of station separation.

Figure 2. Family of synthetic correlation functions.  
(Not to same scale as Figure 1.)
Figure 1 shows an estimated 500 mb correlation function from LH. Let \( h^o_l \) and \( h^f_l \) be the observation and the forecast 500 mb height at station \( l \), and \( \varepsilon^o_l \) and \( \varepsilon^f_l \) be the observation and forecast errors, (on a particular day). Let

\[
\xi_l = h^o_l - h^f_l = \varepsilon^o_l - \varepsilon^f_l
\]

Here, \( y^{(1)} = h^o \) and \( y^{(2)} = h^f \), thus both vectors represent the same meteorological quantity, that condition will be relaxed later. Letting \( \tau_{lm} \) be the distance between stations \( l \) and \( m \), LH assumed that

\[
E \varepsilon^o_l \varepsilon^o_m = \sigma^2_{\xi} \rho(\tau_{lm})
\]

where \( \rho(0) = 1 \). If the \( \varepsilon^o_l \) are independent and identically distributed zero mean random variables then

\[
E \xi_l \xi_m = \sigma_{\xi}^2 \delta_{lm} + \sigma_{\xi}^2 \rho(\tau_{lm})
\]

where \( \delta_{lm} = 1 \) if \( l = m \) and 0 otherwise. LH collected 90 days of values of \( \xi_l \) for each station. Letting \( j \) index day, they used sample correlations computed from

\[
\frac{1}{90} \sum_{j=1}^{90} \xi_l(j) \xi_m(j)
\]

as an estimate of \( \sigma_{\xi}^2 \delta_{lm} + \sigma_{\xi}^2 \rho(\tau_{lm}) \). In the figure, sample values of \( \frac{\sigma_{\xi}^2 \rho(\tau)}{\sigma_{\xi}^2 + \sigma_{\xi}^2 \rho(0)} \) are plotted, and \( \frac{\sigma_{\xi}^2}{\sigma_{\xi}^2 + \sigma_{\xi}^2} \) is estimated by extrapolating the smooth part of the curve back to the origin by methods described in their paper (quite different than the methods to be discussed here.)

Figure 2 shows a one parameter family of (synthetic) correlation functions, defined by

\[
\rho_L(\tau) = \frac{(1-2\theta(L)\cos(\frac{2\pi \tau}{R_O}) + \theta^2(L))^{-1/2} - (1+\theta(L))^{-1}}{(1-\theta(L))^{-1} - (1+\theta(L))^{-1}}
\]

where \( \theta(L) \) is determined by

\[
\frac{3}{2} \theta(L) - \frac{1}{2} \theta^3(L) = \cos \frac{2\pi L}{R_O}.
\]

Here the parameter \( L \) in km is the distance \( \tau \) for which \( \rho_L(\tau) = \frac{1}{2} \rho_L(0) = \frac{1}{2} \), where \( R_O \) is the circumference of the earth, in km. This family of (isotropic) correlation functions on the sphere has been chosen here partly because of a superficial resemblance to some of the curves obtained by LH and partly for mathematical convenience. We wish to use this family as a moderately realistic example of the estimation of a single (important) parameter in \( Q_f \), namely, the correlation half-distance. Further study is needed to determine if it is appropriate to include other factors, such as anisotropy, variation with latitude, etc. in this correlation function. With this model, letting \( \xi = (\xi_1, \ldots, \xi_n) \) be a vector of one day's data, we have that the covariance matrix of \( \xi \) is \( \sigma^2 (I + rQ_f(L)) \), where the \( ln \)th entry of \( Q_f(L) \) is \( \rho_L(\tau_{lm}) \). The GCV estimate of \( \sigma^2 \) and \( L \) can be shown to be the values of \( r \) and \( L \) which minimize
\[ V(r, L) = \frac{\xi \cdot (I + rQ_f(L))^{-2} \xi}{\frac{1}{n} \text{Trace}(I + rQ_f(L))^{-1}^2} \]

If \( \xi \) is assumed to have a multivariate Normal distribution with zero mean, then the maximum likelihood estimates of \( r \) and \( L \) are the minimizers of

\[ M(r, L) = \frac{\xi \cdot (I + rQ_f(L))^{-1} \xi}{\left[ \det(I + rQ_f(L)) \right]^n} \]

Properties of these estimates are under study, and properties and derivations will appear elsewhere. While the ML estimate has various optimality properties if all of the assumptions of the model are satisfied, the GCV estimate may be more robust to model errors, see Seaman and Hutchinson (1985), Wahba (1985b). These remarks are conjectural at the present time.

3.2 WHEN ARE GOOD ESTIMATES POSSIBLE?

We would like to know whether or not it is reasonable to attempt to obtain an estimate of \( r \) and \( L \) from the data to be analyzed. In particular, as \( n \) becomes arbitrarily large, can we expect that the mean square errors of the estimates will become arbitrarily small?

Fortunately, this question can be answered, by applying the mathematical theory of equivalence and perpendicularity. We first state a rather abstract theorem, adapted from Parzen (1963), then we apply it to our example, and make a few remarks concerning its intuitive meaning.

Theorem:

For each \( n \), let \( x = (x_1, \ldots, x_n) \) be a zero mean Normally distributed random vector with covariance matrix \( C = C(\beta) \) where \( \beta \) are some unknown constants. It is desired to estimate \( \beta \) from \( x \). For any two distinct values \( \beta^{(1)} \) and \( \beta^{(2)} \) of \( \beta \), let

\[ J_n(\beta^{(1)}, \beta^{(2)}) = \text{Trace} \left[ (I - C(\beta^{(1)})^{-1/2} C(\beta^{(2)}) C(\beta^{(1)})^{-1/2}) \right]^2. \]

If \( J_n(\beta^{(1)}, \beta^{(2)}) \) tends to infinity as \( n \) tends to infinity for any two distinct possible pairs \( (\beta^{(1)}, \beta^{(2)}) \), then there exists a sequence of estimates \( \beta_n, n = 1, 2, \ldots \) of \( \beta \) whose mean square error goes to zero as \( n \) tends to infinity. If \( J_n \) is bounded as \( n \) tends to infinity for all possible pairs, then there cannot exist such a sequence of estimates.

In the example here, \( C = \sigma^2_I(I + rQ_f(L)) \) and \( \beta = (\sigma^2_I, r, L) \). It can be shown, that if \( \sigma^2_I \) and \( r \) and \( L \) are all restricted to be strictly positive, then \( J_n \) tends to infinity, and good estimates of these quantities are possible (at least in theory), from one data set, if \( n \) is sufficiently large. A deeper analysis of this theorem (and common sense!) suggests that the condition for estimability of \( r \) is that \( Q_f(L) \), for any \( L \), has eigenvalues that are very large compared to 1 as well as very small. Intuitively, it means that the forecast error has a substantial "low frequency" component.

3.3 THE GENERAL CASE
We now consider the general (linear) case, where
\[ y^{(i)} = F^{(i)} s + \varepsilon^{(i)}, \quad i = 1, 2 \]
where \( y^{(i)} \) is of dimension \( n^{(i)}, i = 1, 2 \) and the covariance of \( \varepsilon^{(i)} \) is \( \sigma_i^2 Q_i \). To estimate \( r \), we must be able to construct two matrices \( B^{(1)} \) and \( B^{(2)} \) of dimension \( n \times n^{(1)} \) and \( n \times n^{(2)} \) respectively, which satisfy \( B^{(1)} F^{(1)} = B^{(2)} F^{(2)} \). Let \( z^{(i)} \) be defined by
\[ z^{(i)} = B^{(i)} y^{(i)}, \quad i = 1, 2 \]
and let \( w \) be defined by
\[ w = z^{(1)} - z^{(2)} = B^{(1)} y^{(1)} - B^{(2)} y^{(2)} = B^{(1)} \varepsilon^{(1)} - B^{(2)} \varepsilon^{(2)}. \]
The covariance matrix of \( w \) is then
\[ E w' w = \sigma_1^2 B^{(1)} Q_1 B^{(1)}' + \sigma_2^2 B^{(2)} Q_2 B^{(2)}'. \]
Suppose \( B^{(1)} Q_1 B^{(1)'} \) is of full rank, then we can take the Cholesky decomposition \( LL' \) of \( B^{(1)} Q_1 B^{(1)} \), where \( L \) is lower triangular, and let \( \xi = L^{-1} w \). Then the covariance matrix of \( \xi \) is
\[ E \xi' \xi = \sigma_i^2 (I + rQ), \]
where \( Q = \xi L^{-1} B^{(2)} Q_2 B^{(2)'} L^{-1}' \). The ML and GCV estimates are then given by the minimizers of \( V \) and \( M \) of Section 3.1, and the estimability of \( r \) depends on the properties of \( Q \). Loosely speaking, the two vectors \( z^{(1)} \) and \( z^{(2)} \) need have to have their "energy" at different "wavenumbers".

3.4 COMPUTATIONAL CONSIDERATIONS

Computation of \( r \) and \( L \) minimizing quantities similar to \( M \) and \( V \) has been carried out in a relatively straightforward way using matrix decompositions for \( n \) up to a few hundred, on a VAX 11/720 in the Statistics Dept. at the University of Wisconsin-Madison, and with \( n \) up to about 800 on the Cray XMP at the San Diego Super Computer Center, without much optimization of the code, in under 150 seconds, using GCVPACK (Bates et al. (1985)). Research is continuing on more efficient methods for larger problems. It is probably true that large data sets will be required in practice.

3.5. POTENTIAL APPLICATIONS

A possible important application is to the comparison of satellite observed radiances to forecast. It is possible to observe certain gross features of the atmosphere in two dimensional plots of satellite (raw) radiances. Forecast errors frequently tend to be phase errors, with certain spatial features displaced in space. To compare forecast \( y^{(2)} \) with satellite radiances \( y^{(1)} \) following the approach in this paper, one should compute from the forecast, radiances that would be seen by the satellite, that is, let \( B^{(1)} \) be \( I \) and \( B^{(2)} \) map forecast into radiance data as would be seen by the satellite. Assuming that realistic \( Q_1 \) and \( Q_2 \) can be established (remember, \( Q_2 \) is in the radiances observational domain), it is likely that \( r \) could be estimated, and used to help decide whether to trust the forecast or the radiances data in the event of a major discrepancy.
4. SUMMARY

A study has been initiated into the estimation from the data to be analyzed, of relative weight to be given to various sources of data, for the purposes of numerical weather prediction. For the estimation of these weights to be successful the various sources have to have sufficiently different error correlation structure. The meaning of "sufficiently different" has been quantified in a theoretical way, and reduction of this theory to practical cases has begun with a study of 500 mb height forecast and observational data. Estimates are proposed and numerical methods for computing them for very large data sets are under study. The method has the potential for determining the relative credence to be given to different sources of information, such as forecast and satellite radiance data.

ACKNOWLEDGMENT

This work has benefitted from numerous helpful discussions with D. R. Johnson.

BIBLIOGRAPHY


O'Sullivan, F. and G. Wahba (1985), A cross validated Bayesian retrieval algorithm for non-


Hampton, VA.
