4.3

A NEW APPROACH TO THE OPTIMAL ASSIMILATION OF METEOROLOGICAL DATA BY ITERATIVE BAYESIAN ANALYSIS

R. James Purser
Cooperative Institute for Meteorological Satellite Studies
(On leave from United Kingdom Meteorological Office)
University of Wisconsin
1225 West Dayton Street
Madison, Wisconsin 53706

1. INTRODUCTION

The requirement of modern objective data assimilation schemes to accommodate data of different types and variable reliability has led to an increasing emphasis on the use of statistically based methods of analysis in the preparation of initial fields for numerical prediction models. Notable contributions to the development of analysis schemes that are in some statistical sense optimal have been made by Sasaki (1958) and by Gandin (1963) who pioneered the technique of optimum interpolation. Developments by Rutherford (1972), Schlatter (1975), Bergman (1979), and Lorenc (1981) have established it as an effective practical technique.

This paper describes a method of analysis currently under development which generalizes the linear formulation of optimum interpolation to an essentially non-linear one, exploiting a statistical approach based on Bayes' theorem of conditional probabilities. While under special restricted conditions the Bayesian approach becomes identical to the linear optimum interpolation, a more general non-linear formulation appears formally to be able to handle in a statistically consistent and unified way several aspects of the data assimilation problem that have hitherto been dealt with separately. These include the problem of data "quality control," i.e., how to treat the occasional, but potentially damaging, occurrence of observations that, for unknown reasons, possess abnormally large errors; the consistent inclusion of non-linear balancing procedures or "initialization" and the direct insertion of satellite derived data avoiding the separate intermediate step of performing independent single-column retrievals of temperature and humidity. In addition, this formulation can accommodate a number of adaptive features that were sometimes present in empirical analysis methods, such as the stretching or bending of structure functions of "successive correction" schemes according to the local flow features that are intuitively desirable but which are lacking in conventional optimum interpolation.

The Bayesian scheme is presented in relation to conventional optimum interpolation. Being non-linear it demands the use of iterative methods and the recognition of this fact has strongly influenced the composition of the algorithms designed to achieve the desired optimal analysis. Consideration has been given to strategies that avoid where possible the explicit manipulation (e.g., inversion) of very large matrices which would consume an inordinate amount of computation, and an outline is sketched of the algorithmic structure developed to attain this objective.

2. OPTIMUM INTERPOLATION AND BAYESIAN GENERALIZATIONS

In optimum interpolation an analysis A_i consisting of one or more variables at each gridpoint i, is obtained as a linear combination of a "background" field B_i (also known at each gridpoint) and an incomplete scatter of observations O_α located at observations points α. The linear coefficients are chosen to minimize the expected mean square of the analysis error A_i' and are derived from knowledge or estimates of the covariances C_{ij} of background field errors B_i' and from the covariances $E_{\alpha\beta}$ of observational errors O_α', assuming that all errors are unbiased and that the set $\{O_\alpha'\}$ are uncorrelated with set $\{B_i'\}$. It is convenient to use i,j,k, . . . to label standard gridpoint values and α,β,γ, . . . to denote individual observables. Also a function $D_\alpha(\underset{\sim}{B})$ will be used to express the composition of observable α (e.g., a particular satellite radiance observation) in terms of the standard gridpoint values. With these conventions and assumptions it will be stated without proof that the analysis sought is given by the matrix equation

$$A_i = B_i + \sum_{\alpha\beta} (CD)_{i\alpha} (D^T CD+E)^{-1}_{\alpha\beta} (O-D(B))_\beta \qquad (1)$$

where

$$D_{i\alpha}(\underset{\sim}{B}) \equiv \frac{\partial D_\alpha(\underset{\sim}{B})}{\partial B_i} \text{ and } D^T_{\alpha i} \equiv D_{i\alpha} .$$

Using the above subscript convention no confusion will arise by identifying

$$C_{i\alpha} \equiv \sum_j C_{ij} D_{j\alpha}$$

$$C_{\alpha\beta} \equiv \sum_{ij} D^T_{\alpha i} C_{ij} D_{j\beta}$$

$$A_\alpha \equiv D_\alpha(\underset{\sim}{A}) \cong D_\alpha(\underset{\sim}{B}) + \sum_i (A_i - B_i) D_{i\alpha}$$

A significant feature of (1) that is immediately evident is that a matrix inverse is required, the order of the matrix being formally equal to the number of observations considered. In a large scale analysis system based on the Gandin method it is of course impossible to solve the formal system (1) and it is customary to restrict severely the number of observations permitted to influence each grid point.

An alternative approach to analysis optimization can be developed from probabilistic principles. Imagine a "state" of the system, i.e., its N gridpoint values, as represented by a point in an N-dimensional space whose coordinates are the possible gridpoint values themselves, thus both the analysis $\underset{\sim}{A}$ and background $\underset{\sim}{B}$ can be regarded as position vectors in this "state-space." Similarly the M observables may be thought of as defining a point in an M-dimensional "observation space." Each individual observable α is associated with a continuous family of (N-1)-dimensional surfaces in state-space parametrized naturally by the values of D_α. In practice, knowledge of the state is always somewhat vague and may be formally regarded at any time as a probability density function in state-space, for example, the prior knowledge or assumptions of the "location" of the atmospheric state are summarized by a probability density, $P_B(\underset{\sim}{B}')$ of the errors $\underset{\sim}{B}'$ of the initial guess $\underset{\sim}{B}$ comprising the locally most probable state. Note that this choice for $\underset{\sim}{B}$ implies

$$\left.\frac{\partial P_B(\underset{\sim}{B}')}{\partial B_i'}\right|_{\underset{\sim}{B}'=0} = 0 . \qquad (2)$$

Similarly, the observations are known to contain random errors or to be contaminated by effects too small or too transient to be significant so it is natural to express this degree of vagueness also in probabilistic terms. For example, by assuming that the observation errors O_α' are each distributed independently by a probability density $P_\alpha(O_\alpha')$. Assuming the prior distribution P_B is obtained independently from the new observations (e.g., using climatology, a previous forecast together with dynamical constraints) then it is possible to combine the two sources of information into a single conditional probability distribution, say P_A, using Bayes' rule for conditional probabilities.

$$P_A(\underset{\sim}{A})=\rho P_B(\underset{\sim}{B}-\underset{\sim}{A}) \prod_{\alpha=1}^{M} P_\alpha(O_\alpha-A_\alpha) \qquad (3)$$

where ρ is a normalizing factor. Figure 1 illustrates schematically the typical application

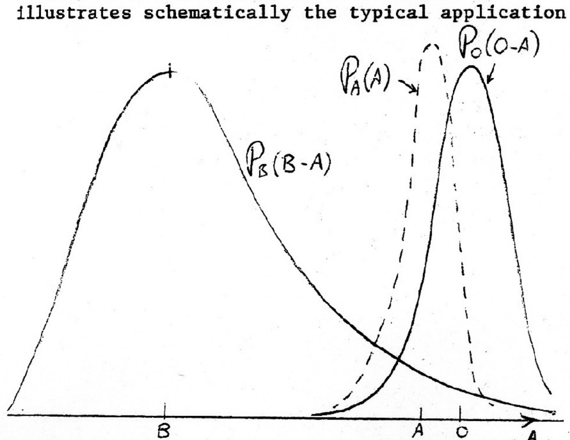

Figure 1: Schematic illustration of Bayes' rule in one dimension.

of this rule to one dimension, the conditional probability density P_A being peaked between the peaks of distributions P_B and P_O, to be narrower than either of them and to be closer to the narrower distribution P_O than to P_B. Following the spirit of Gandin's optimum interpolation one would consider the centroid of P_A as the optimal analysis since only at this location in state-space is the expectation squared error of each component of A_i of the analysis simultaneously minimized. However, a simpler procedure, though one that is arguably less "optimal," is to redefine the optimal analysis as that which <u>maximizes</u> P_A itself (or equivalently, its logarithm) with respect to local variations of the components A_i, i.e.,

$$\frac{\partial}{\partial A_i}\log P_A \equiv \frac{\partial}{\partial B_i^*}\log P_B(\underset{\sim}{B}^*)+\sum_{\alpha=1}^{M} D_{i\alpha}\frac{\partial}{\partial O_\alpha^*}\log P_\alpha(O_\alpha^*)=0 \qquad (4)$$

where

$$O_\alpha^*=O_\alpha-A_\alpha$$

$$B_i^*=B_i-A_i \, .$$

This general formulation can be shown to reduce to standard linear optimum interpolation with independent observational errors under the assumptions of Gaussian structures to P_B and P_α and the linearity of function D.

Then

$$P_B(\underset{\sim}{B}') \propto \exp\{-\frac{1}{2}\sum_{ij} C_{ij}^{-1}B_i'B_j'\}$$

$$P_\alpha(O_\alpha') \propto \exp\{-\frac{1}{2}E_\alpha^{-1}O_\alpha'^2\} \qquad (5)$$

reducing (4) to

$$\sum_{ij} C_{ij}^{-1}B_j^* + \sum D_{i\alpha}E_\alpha^{-1}O_\alpha^*=0$$

hence

$$A_i = B_i + \sum C_{ij}D_{j\alpha}E_\alpha^{-1}(O_\alpha-A_\alpha) \qquad (6)$$

which is equivalent to (1) when A_α is eliminated from the right side.

Under more general circumstances "effective" statistics can be obtained to replace E_α^{-1} and C_{ij}. In principle these are obtained from the local behavior of the probability densities P_B and P_α in the vicinity of B^* and O^* and formally yield pseudo-covariance C_{ij} whose matrix inverse is

$$C_{ij}^{-1} = -\{\frac{\partial}{\partial B_i^*}\log P_B(B^*)\}\frac{B_j^*}{\sum_k (B_k^*B_k^*)} \qquad (7)$$

and the pseudo-variance E_α whose inverse is

$$E_\alpha^{-1} = -\frac{\partial}{\partial O_\alpha^*}\,\frac{\log P_\alpha(O_\alpha^*)}{O_\alpha^*}\,. \qquad (8)$$

An important point to make here is not that one should attempt to catalogue a comprehensive tabulation of probabilities P_β and P_0 and extract effective statistics as in (7) but rather that one should recognize that the Bayesian method is flexible enough to accommodate effective statistics that are adaptive to the situation being analyzed and that it provides guidance as to how this might be done.

A suggestion of the versatility and potential of the adaptive formalism is provided by illustrating its handling of observations known occasionally to contain gross errors, i.e., those for which a traditional analysis method requires a separate quality control procedure. For simplicity suppose the errors of observable α have a probability density:

$$P_\alpha(O_\alpha') \propto \exp\{E_0^{-1}\, a\, [\exp(-\frac{O^{*2}}{2a}) - 1]\} \qquad (9)$$

as illustrated in Figure 2a for $a=2E_0=2$. Note the existence of non-vanishing tails in this distribution, consistent with the occasional appearance of gross errors. The impact of such an observable on the analysis as a function of its final departure from that analysis is given by the "forcing function:"

$$X_\alpha(O_\alpha^*) \equiv E^{-1}O_\alpha^* = -\frac{\partial}{\partial O_\alpha^*}\log P_\alpha(O_\alpha^*)$$

i.e.,

$$X_\alpha(O_\alpha^*) = E_0^{-1}\exp(-\frac{O^{*2}}{2a})\,O_\alpha^* \qquad (10)$$

(Figure 2b). It is evident that the impact of the observation is negligible if a large disparity persists between it and the analysis. To complete the picture (Figure 2c) the effective weight is, by (8):

$$E^{-1}(O_\alpha^*) = E_0^{-1}\exp(-\frac{O_\alpha^{*2}}{2a})\,. \qquad (11)$$

This artificial example demonstrates the ability of the formalism to incorporate in a natural way a form of quality control, but by continuous weighting rather than by an explicit rejection-acceptance criterion.

3. COMPUTATIONAL CONSIDERATIONS

The implicit optimal analysis equation (6) forms the core of the iterative methods. Since it contains no non-trivial matrix inverse it is simple to verify. Suppose an approximation A to the optimal analysis is obtained by replacing $E_\alpha^{-1}(O_\alpha - A_\alpha)$ in (6) by an approximation to the forcing, X_α. The degree of inconsistency between

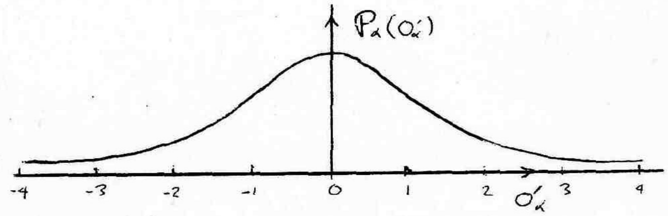

Figure 2a: P_α as a function of O_α'.

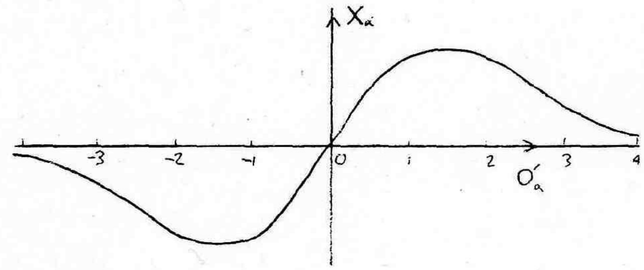

Figure 2b: X_α as a function of O_α'.

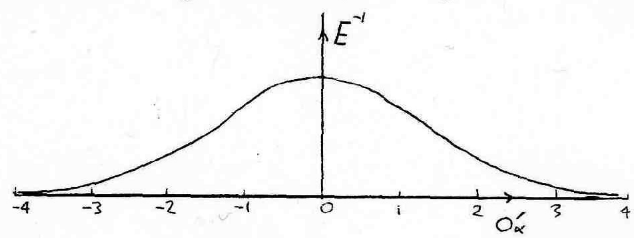

Figure 2c: E_α^{-1} as a function of O_α'.

these approximations is measured by evaluating the residual,

$$R_\alpha = E_\alpha^{-1}(O_\alpha - A_\alpha) - X_\alpha \qquad (12)$$

for each observable (only when the residual vanishes for every observable is the analysis obtained in this way "optimal"). The sensitivity of $\underset{\sim}{R}$ to changes in X is

$$\frac{\partial R_\alpha}{\partial X_\beta} = -(E_\alpha^{-1}C_{\alpha\beta} + I_{\alpha\beta})$$

where $I_{\alpha\beta}$ is the unit matrix. Then provided an approximation to the inverse of this matrix is available, one may obtain an improved estimate of X and hence of A by means of the correction:

$$X_\alpha = X_\alpha + \sum_\beta G_{\alpha\beta}R_\beta \qquad (13)$$

where

$$G_{\alpha\beta} \cong (E^{-1}C+I)^{-1}\,.$$

The repetition of this procedure will lead to successively better analyses, $\underset{\sim}{A}$. Adaptive features of the Bayesian method discussed earlier, i.e., the updating of C, D, E^{-1}, may be included. In addition, it becomes feasible to incorporate periodically adjustments that insure that a state of dynamical balance is maintained, thereby combining "analysis" and "initialization" in a single scheme. Briefly, this is achieved by inserting the unbalanced analysis, now denoted $\underset{\sim}{\hat{A}}$ into the forecast model which is then run forward for two timesteps. The first and second time derivatives of divergence provide convenient diagnostics of dynamic imbalance which can be corrected by applying these diagnostic fields as forcings to equations analogous to the balance and omega equations to obtain correction fields that bring the analysis back towards a state of balance, $\underset{\sim}{A}$. Symbolically this procedure is written,

$$A_\alpha = J_\alpha(\underset{\sim}{\hat{A}})$$

and the iterative scheme I have described then matches the flow diagram, Figure 3.

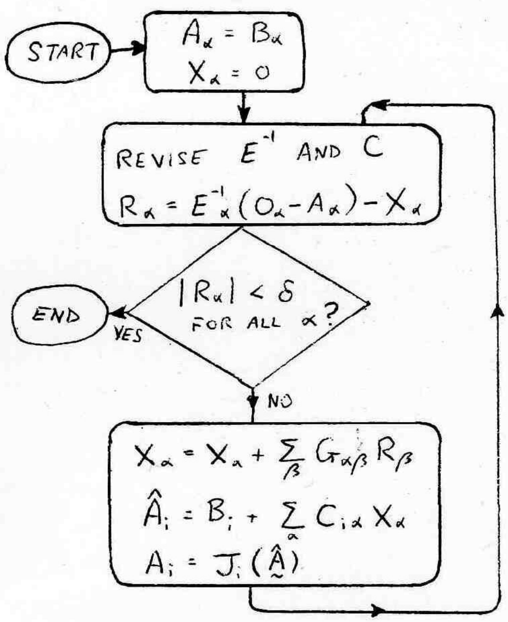

Figure 3: Flow chart for an iterative Bayesian analysis algorithm.

A major difficulty with this algorithm is obtaining an estimate for $G_{\alpha\beta}$ which is good enough to give a reasonably rapid convergence of the algorithm yet which does not require the direct computation of large matrix inverses. There are indications that the problem is alleviated by using a method able to discriminate between the different spatial scales at which analysis corrections are required and to treat these scales separately. One approach is to organize the observations into small clusters and to exploit the fact that the forcing, X_β formed by the sum of forcings X_α in a tight cluster β and acting at the centroid of this cluster, has an impact almost identical to that of the original forcings X_α. In this way it is possible to approximate the analysis problem by a coarser-scale representation containing fewer elements. The corrections deduced at the coarser scale are enforced as temporary strong constraints of a modified analysis problem when the iterative algorithm descends back to the finer scale. The clustering procedure, and the associated adjustments to the analysis algorithm are readily extended to a hierarchy of levels of clustering. It is found that reliably good approximations $G_{\alpha\beta}$ to the correction matrices at each level of the hierarchy may now be obtained without excessive computation. This strategy is analogous to the "multigrid" algorithm developed by Brandt (1977) and others to accelerate the convergence of iterative solutions of elliptic equations.

4. CONCLUSION

An alternative statistical formalism for dealing with general problems in meteorological data assimilation has been introduced. This enables several aspects of the problem previously treated separately, notably those implicitly requiring nonlinear treatments, to be brought together within a unified framework. Methods of solving the resulting equations are necessarily iterative and this fact is used to advantage in the design of algorithms that avoid the direct and costly inversion of large matrix systems which appear in conventional optimum interpolation. By transforming the problem into a series of representations at progressively coarser scales it is possible to enhance the efficiency of the algorithm as a whole. A high-priority application of the technique is to the three dimensional assimilation of satellite radiances to circumvent the somewhat dubious custom of treating retrieved satellite temperatures in an analysis as if they comprise a set of observations with errors independent of those of the background field. If successful a unified and consistent approach to the analysis of mixed satellite and ground based observation would be of obvious benefit.

5. REFERENCES

Bergman, K. H., 1979: Multivariate analysis of temperatures and winds using optimum interpolation. Mon. Wea. Rev., 107, 1423-1443.

Brandt, A., 1977: Multi-level adaptive solutions to boundary-value problems. Math. Comp., 31, 333-390.

Gandin, L. S., 1963: Objective analysis of meteorological fields. Leningrad, Gidromet; (Jerusalem, Israel Program for Scientific Translations; 1965, 242 pp.).

Lorenc, A., 1981: A global three dimensional multivariate statistical interpolation scheme. Mon. Wea. Rev., 101, 701-721.

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. J. Atmos. Sci., 29, 809-815.

Sasaki, Y., 1958: An objective analysis based on the variational method. J. Meteor. Soc. Japan, 36, 77-88.

Schlatter, T. W., 1975: Some experiments with a multivariate statistical objective analysis scheme. Mon. Wea. Rev., 103, 246-257.