Tellus (1982), *34*, 113–126

Stochastic simulation of the atmosphere with localized precision

By D. K. PURNELL, *New Zealand Meteorological Service, P.O. Box 722, Wellington 1, New Zealand*

(Manuscript received June 17, 1980; in final form April 14, 1981)

ABSTRACT

In the conventional representation of the distribution of error in an estimate of the state of the atmosphere it is usual to assume, for the sake of computational economy, that error covariance is localized. There is less justification for this assumption than appears to be commonly supposed. For instance, an array of observations of wind, whose observational errors are independent of the "first guess" error and independent of each other, will actually spread the covariance of error in geopotential arbitrarily far, and increase the correlation of error in the geopotential to a maximum as the error in the wind observations tends toward zero. On the other hand, localized independent observations do not spread precision, which is the inverse of covariance. The coupling between wind and geopotential can be described by localized precision, so that there is the possibility of describing the entire problem in this way.

The stochastic-dynamic simulation problem is formulated in terms of precision, with a linearizing approximation for the evolution of small errors. It is concluded that the precision will be approximately localized if the simulation is driven by a sufficient amount of localized observations. A method of computing stochastic simulations is devised which allows drastic computational economies when precision is localized.

The precision formulation is demonstrated by experiments with a simple stationary model.

1. Introduction

Much of the development of variational methods for analysis and initialization of meteorological data is attributable to Sasaki (1970a, b, c). An aspect which he seems to have omitted is the interpretation of his "variational formalism with weak constraint" (Sasaki, 1970a) as a "least-squares" estimator in a stochastic setting. The stochastic viewpoint is natural, considering that very little is known exactly in any simulation of the atmosphere, and that the theory of statistical estimation provides a logical basis for analysis of conflicting data.

From this point of view, the difficulties of estimating the state of the atmosphere from measurements are first, that of formulating an adequate but practical representation of the probability distribution of errors in the data and in any prior information on the state of the atmosphere, and second, that of computing an estimate with desirable statistical properties from the data. These two aspects have not always been explicitly recognized in analysis methods, but some model of the error distribution is always implied.

Estimates of the state of the atmosphere are not in practice based solely on contemporaneous data. Optimal estimation from data distributed over time will require consideration of the time-evolution of the error distribution of the estimate as in Epstein and Pitcher (1972) and Pitcher (1977), but in the method to be proposed for the calculation of short-term evolution there is much more scope for a compromise between ideals of generality, accuracy, and computational economy. This is to be based on two main ideas, the idea of a spatial localization of precision, and the idea of a linear approximation for the evolution of small errors. This linearization gives the deterministic approximation of the mean, but it gives the same evolution of covariance as the linearization proposed by Laurmann (1978). In his experiments, Laurmann

0040-2826/82/020113-14$02.50/0

found "no reason for not using the linearized approximation". It would be possible to compute Laurmann's correction to the mean from covariances computed by the method proposed in the present paper. It is suggested however that for many applications this would be unnecessary, since data assimilation will require only short-period predictions, and the present linearization can be expected to be accurate until uncertainty becomes large, when a user of forecasts is likely to have lost interest anyway.

2. Least-squares estimation

2.1. *The general model*

Suppose that h is a finite-dimensional representation of a state of the atmosphere. In practice this is a representation of winds, temperatures, etc., either by grid-point values, with an associated interpolation rule, or else by coefficients of a finite set of basis functions. In either case it is possible to assign a unique representation h to any particular state of the atmosphere, for example, in the Galerkin procedure we choose that representation which is nearest to this state in the sense of a suitable norm. In particular, there is a unique representation z corresponding to the true state of the atmosphere at a particular time. The problem which is to be considered is that of estimating z from knowledge of the data d, which consists of a list of measurements, and from knowledge of a "first guess" g, which is an estimate of z based on previous data.

The point of view adopted is that d, g and z are sampled from a probability distribution with density

$$\begin{aligned}\mu(d, g, z) &= \mu_z(d, g)\rho(z) \\ &= \phi_z(d)\theta_z(g)\rho(z)\end{aligned} \tag{1}$$

where μ_z is a density conditional on the realization of a particular state z, and error in the data is assumed to be distributed independently of error in the guess. It is now assumed that the three independent distributions

$$\phi_z(d), \quad \theta_z(g), \quad \rho(z)$$

are generated by respective subsets of the regression model

$$\begin{aligned}R(d - Dz) &= e_d \\ A(g - z) + a &= e_g \\ C(c - z) &= e_c\end{aligned} \tag{2}$$

where e_d, e_g, e_c are random variables to account for error in the data and in the guess, and to account for the difference between the true state and climatological mean state c. The operators R, A, C and D, and the bias a are to be chosen so that the components of e_d, e_g, e_c are distributed independently with zero expectation and unit variance. Then the least-squares estimate of z for this regression problem is h, where h minimizes

$$\begin{aligned}\xi = &\|R(d - Dh)\|^2 + \|A(g - h) + a\|^2 \\ &+ \|C(c - h)\|^2\end{aligned} \tag{3}$$

where $\|\cdot\|$ is the Euclidean (root-mean-square norm on the respective implied spaces.

For most of the discussion it will be sufficient to consider the case where the operators R, A, C and D are linear, and it will be convenient to use their matrix representations. No confusion should be caused by the subsequent use of the same symbol to represent both a linear operator and its matrix. The transpose and inverse of a matrix S will be denoted by S' and S^{-1} respectively.

From eq. (2), the covariance of e_c is

$$I = \text{expectation } (e_c e_c') = CHC'$$

where I is the identity and H is the covariance of the distribution of z about the climatological mean. Since H will be positive definite a suitable choice for C would be $C = C' = H^{-1/2}$. Similarly, the covariance V of error in the data is positive definite (assuming there is no perfect data) and for the following take $R = R' = V^{-1/2}$. If there is no correlation of error between separate items of data then R is block-diagonal, one block for each item. A construction for A will be developed on the following lines, in the context of a time series of estimations of z. The guess for each estimation will be computed from the previous estimate, so that A will be updated at each step. It will be shown (eq. (13)) that a suitable update can be obtained simply by appending the rows of RD to the rows of A. Furthermore, in a time series of such estimations, A will consist entirely of such appended rows since there will turn out to be good reason for deleting old rows.

The operator D will be called "the observation

operator" for the data d, meaning that Dz is an unbiased estimate of d, given z. In practice D will be an interpolation of limited resolution and interpolation error will be lumped together with error in the data measurements in all subsequent discussion.

The operators R, A and C have the role of weighting functions for the data, guess, and the fit to climatology. In the absence of data this implies a regression back to climatology as the uncertainty in the guess increases toward the point where C dominates A.

Equations (2) and (3) incorporate climatology into the model, but it will be simpler to restrict attention to the main problem, which is that of modelling the conditional probability distribution $\mu_z(d, g)$, the distribution of error about the true state at a particular time. For this eq. (3) reduces to

$$\xi = \|R(d - Dh)\|^2 + \|A(g - h) + a\|^2 \tag{4}$$

Then the solution minimizing ξ is h where

$$D'R'R(Dh - d) + A'A(h - g) = A'a \tag{5}$$

or

$$h = g + (D'R'RD + A'A)^{-1}(D'R'R(d - Dg) + A'a) \tag{6}$$

The precision matrix is defined to be the inverse of the covariance matrix. Then from eq. (2)

$$I = \text{expectation } (e_g e_g') = AQA'$$

where Q is the covariance of the guess error. So the precision of the guess g is $Q^{-1} = A'A$, and similarly, the precision of the data d is $R'R$. From eq. (5) the estimate h is

$$h = (D'R'RD + A'A)^{-1}(D'R'Rd + A'Ag + A'a)$$

so that, substituting the guess and data covariances,

$$\text{covariance } (h) = \text{expectation } ((h - \text{expectation } (h)) \times (h - \text{expectation } (h))') = (D'R'RD + A'A)^{-1}$$

and

$$\text{precision } (h) = D'R'RD + A'A. \tag{7}$$

This is just the sum of the precisions contributed by the guess and data.

Of all linear unbiased estimates, the least-square estimate is the one with minimum variance, but it may have disadvantages in other respects if the distributions of e_d, e_g, e_c are not normal. However, if the fields of interest are smooth and error is small, then a linear approximation may be used for the evolution of error, and approximate normality will be guaranteed by the central limit theorem, since the data analyses have the character of an average of a large number of independent data items.

2.2. *Relation to other estimation schemes*

Many analysis schemes (e.g. Cressman, 1959) have been constructed using weighting functions for the observations, and the second term in eq. (6) can be interpreted in this way. Since in eq. (6) h is a linear function of the observations d, the principle of superposition of the influences of individual observations applies and the weighting function for the ith observation is the ith column of the matrix

$$W = (D'R'RD + A'A)^{-1} D'R'R \tag{8}$$

which operates on the difference $(d - Dg)_i$ between the observation and its "first guess' estimated value. These weighting functions vary with the type and precision of nearby data. They also depend on the guess precision through the term $A'A$.

Equation (8) can be shown to be equivalent in principle to a generalized form of the widely used "optimum interpolation" method of Gandin (1963), as follows. With the above notation we define an estimate $\hat{z}$ of z to be

$$\hat{z} = g + W(d - Dg) \tag{9}$$

where W is a matrix of weights. The Gandin (1963) method essentially is a matter of choosing W so as to minimize the estimation error variance, and we now show that this choice is given by eq. (8), given the assumptions of Section 2.1. The ith component of the estimation error is

$$e_i = ((z - g) - W(d - Dg))_i.$$

If g has covariance Q and d has covariance V, the estimation error variance is

$$\sigma_i^2 = \text{expectation } (e_i^2) = ((WD - I)Q(D'W' - I) + WVW')_{ii}$$

where I is the identity matrix. This variance is minimized when

$$0 = \partial\sigma_i/\partial W_{ij}$$
$$= (W(DQD' + V) - QD')_{ij}$$

or

$$W' = (DQD' + V)^{-1} DQ. \tag{10}$$

The usual "optimal interpolation" methods appear to follow eqs. (9), (10), except that for reasons of economy, an additional hypothesis is made that W, $DQD' + V$, and DQ are sparse; both the weighting functions and the covariance are assumed to be localized. If Q and V are positive definite then from eq. (10)

$$W = QD'((DD')^{-1} D(D'V^{-1} D + Q^{-1})QD')^{-1}V^{-1}$$
$$= (D'V^{-1} D + Q^{-1})^{-1}D'V^{-1}$$

which is the same as eq. (8) when we substitute

$$V^{-1} = R'R, \quad Q^{-1} = A'A.$$

Least-squares estimation theory has a natural extension from finite-dimensional to general Hilbert spaces. Such an extension would be an example of "generalized spline smoothing". A two-parameter spline smoothing method for meteorological problems has recently been reported by Wahba and Wendelberger (1980).

3. Covariance structure

3.1. *Efficient representation of covariance*

Estimation schemes for the atmosphere have usually been formulated in terms of covariance, and to achieve computational economy it is usual to make the assumption that covariance of error at widely spaced locations can be neglected. In other words, it is usual to use localized covariance. However, it will be argued that a localized approximation of the precision is more promising.

In the type of problem being considered, the error is an error of prediction and of estimation. The distribution of this error must not be confused with the distribution of states about the climatological mean. One of the main points of this paper is that approximate localization of precision can be justified in problems of this type on the basis of some plausible assumptions. This justification rests on the properties of the least-squares estimate, the nature of the observations, and the general character of the growth of uncertainty. The prototype argument is as follows. Most, if not all, of the measurements that are made are spatially localized, that is to say the observations are not influenced by the state of the atmosphere at distant locations. Suppose the precision of the guess g is localized. If the errors of measurement are uncorrelated (i.e. if R is diagonal), the precision of the estimate h will be localized because any sum of localized precisions is again localized, and the precision of the least-squares estimate is such a sum (see eq. (7)). To summarize, observations of this type will not spread the precision. Finally, we note (Sections 3.3, 3.4) that spreading of precision by the stochastic-dynamics will occur at only a slow rate, that the predictive value of observations declines with age so that the precision contributed by old observations is negligible, and that contributions from new, localized observations will therefore dominate to the extent that the total precision will be approximately localized.

In contrast to this, an hypothesis that covariances remain approximately localized is indefensible, at least without stringent requirements for the observational network, since it will be demonstrated by example that independent localized observations can spread the covariance arbitrarily far, and increase the correlation of error between distant places to an arbitrary degree. This effect will be pronounced wherever there is a "gap" in the observations in the sense that there is insufficient data to estimate some non-localized mode, so that the error variance of this mode becomes prominent.

To present this argument more formally, let D_i be the observation operator for the ith observation. Let δ_j be the column vector in the domain of D_i whose elements are zero, except for the jth, which is unity. Define the "support" of the ith observation to be the set of integers

$$S_i = \{k \mid D_i \delta_k \neq 0\}.$$

In other words, the estimate $D_i z$ depends only on those elements of z which belong to S_i. Define a distance $\Delta(j, k)$ between any pair j, k. For example, in a grid-point representation let $\Delta(j, k)$ be the

number of grid intervals separating the jth and kth elements. Define the "diameter of support" for the ith observation to be

$$r_i = \text{maximum}\,\{\Delta(j,k) \mid j,k \text{ belong to } S_i\}.$$

Then,

$$(D_i' D_i \delta_k)_j \neq 0$$

only if j, k belong to S_i, which implies $\Delta(k,j) \leqslant r_i$. If all observations are local, with diameters $r_i \leqslant r$ say, and R is diagonal, then $(D'R'RD\delta_k)_j \neq 0$ only if $\Delta(k,j) \leqslant r$. Call r the "radius of support" of the matrix $D'R'RD$. Any sum of matrices with radius r, also has radius r, and is obviously sparse if r is small. A representation by grid-points, or by basis functions with localized support, for example, will allow sparse matrices for localized observations.

It is to be noted that the assumption that R is diagonal (or at least local), is not likely to be very restrictive in practice, even if instrumentation errors are not independent. This is because the "observation error", as defined in Section 2.1, includes error of interpolation (i.e. sub-grid scale error). If the instrumentation of observations is more accurate than the interpolation, then the covariance of "observation error" will be dominated by its interpolation component. The scale of these errors is below the spatial resolution of the model, by definition, so that if the observations are separated by distances greater than this scale, it is reasonable to hypothesize that the interpolation errors will be independent, and hence R will be diagonal. If the resolution of the model is sufficient, point-observations will be separated in this way, which will guarantee independence. If not, a separation of groups of observations will make R block-diagonal, by the same argument.

To show that localization of covariance is not preserved by the least-squares estimate, a simple counter-example is sufficient. Consider estimation of a scalar variable Φ, represented on a row of n grid-points, $\Phi_j, j = 0, 1, \ldots, n-1$.

Suppose the guess for Φ has the localized covariance matrix Q, $Q_{j,k} = 0$ except for

$$Q_{j,k} = 1 \text{ if } j = k, \quad Q_{j,k} = \alpha \text{ if } j = (k \pm 1) \bmod (n).$$

Suppose that independent observations of the gradient of Φ are made at $j + \frac{1}{2}, j = 0, 1, \ldots, n-1$, uncorrelated with the guess, with a common observation error variance σ^2. Choose the localized observation centred-difference operator D,

$$D_j(\Phi) = \Phi_{j+1} - \Phi_j, \quad j = 0, 1, \ldots, n-1.$$

Then the precision of the least-squares estimate $\hat{\Phi}$ is $(Q^{-1} + D'D/\sigma^2)$, and the covariance of its error is $\hat{Q} = (Q^{-1} + D'D/\sigma^2)^{-1}$. We note that $\hat{Q}$ is not localized: indeed, the covariance of any pair $\hat{\Phi}_j$, $\hat{\Phi}_k$ is (see appendix)

$$\left(1 + 2\alpha + \sigma^2 \sum_{m=1}^{n-1} \cos(2\pi m(j-k)/n)\lambda_m \right.$$
$$\left. /(\sigma^2 + 2\lambda_m \mu_m)\right)/n$$

where

$$\lambda_m = 1 + 2\alpha \cos(2\pi m/n)$$

$$\mu_m = 1 - \cos(2\pi m/n).$$

Note that the correlation of any pair of elements $\hat{\Phi}_j$, $\hat{\Phi}_k$ tends toward unity as σ^2 tends toward zero. This result is hardly surprising, since the one degree of freedom left when the gradients are known exactly is that of a perturbation which is common to all elements. Finally, note that this example is highly relevant to the effect of an array of wind observations on the covariance of geopotential, since winds are geotropically coupled to gradients of geopotential.

3.2. *Evolution of precision with time*

The covariance of error in the guess depends, of course, on how the guess is produced. It is assumed to be part of the output of a numerical prediction model producing a time-series of guesses, and it is assumed that error in the data forcing this process is serially uncorrelated. Any prediction model can be written

$$x_\tau = P x_{\tau-1} \tag{11}$$

where x_τ is a vector of the meteorological variables, τ is the time in units of one time-step, and P is the (non-linear) operator representing a finite difference approximation to the net tendency in terms of horizontal gradients etc. Associated with eq. (11) is an operator P_τ, which predicts the evolution of small errors Δx,

$$\Delta x_\tau = P_\tau(\Delta x_{\tau-1})$$

where

$$x_\tau + \Delta x_\tau = P(x_{\tau-1} + \Delta x_{\tau-1}).$$

In the following P_τ is assumed to be a linear function of the error. This is a good approximation for small errors provided P is differentiable. It implies that if expectation $(\Delta x_0) = 0$ then expectation $(\Delta x_\tau) = 0$. This approximation seems to be a good one in practice; the effect of non-linearity may be of only academic interest, because non-linear effects become significant only if there is significant probability of an error comparable in magnitude to x. In that case most of the predictive value of the model has evaporated anyhow. This is evident in Leith's (1974) results, from which the average of a number of non-linear predictions appears not much more useful than a single prediction from x_0.

If the true state of the atmosphere has the representation z, and at time τ_{-1} there is an error $x_{\tau-1} - z_{\tau-1}$, with covariance

$$(E_{\tau-1})^{-1} = \text{expectation}\,((x_{\tau-1} - z_{\tau-1}) \times (x_{\tau-1} - z_{\tau-1})')$$

then if no error is introduced by the prediction model, the covariance of $Px_{\tau-1} - z_\tau$ is

$$\text{expectation}\,(P_\tau(x_{\tau-1} - z_{\tau-1})(P_\tau(x_{\tau-1} - z_{\tau-1}))')$$

i.e. the precision of $x_\tau - z_\tau$ is

$$\hat{E}_\tau = (P_\tau^{-1})' E_{\tau-1} P_\tau^{-1}$$

If at time τ there is introduced data of precision $R_\tau' R_\tau$ which is interpolated by D_τ, the precision of the least-squares estimate is (from eq. (7))

$$E_\tau = (P_\tau^{-1})' E_{\tau-1} P_\tau^{-1} + D_\tau' R_\tau' R_\tau D_\tau. \qquad (12)$$

3.3. *Efficient computation of the evolution of precision*

While the evolution can be computed as in eq. (12), a procedure with a different and attractive computational character can be devised for the matrix A of eqs. (2)–(7), where

$$E_\tau = A_\tau' A_\tau$$

so that eq. (12) is satisfied by the choice

$$\mathbf{A}_\tau = \begin{bmatrix} A_{\tau-1} P_\tau^{-1} \\ R_\tau D_\tau \end{bmatrix}.$$

Thus it is possible to do a backward timestep, without consideration of the internal details of the prediction model used, to advance A one timestep as follows. The ith column of $A_{\tau-1} P_\tau^{-1}$ is

$$A_{\tau-1}(P^{-1}(x_\tau + \delta_i) - P^{-1} x_\tau)/\|\delta_i\|$$

where δ_i is the column vector whose elements are zero, except for the ith, which is sufficiently small for the linearizing approximation to be accurate.

The rows of A can be thought of as independent constraints which will apply to subsequent analyses. An update of the ith column of A corresponds to an update of the ith values of all constraints whose domains of influence include the ith variable. It follows that a number of point values can be updated simultaneously, so long as they are separated in the sense that no constraint influences more than one such point. Thus $A_{\tau-1} P_\tau^{-1}$ is computed by n backward timesteps from n slight perturbations of x_τ, where n depends only on the size of the domains of the constraints at τ. Evidently these domains spread at the rate characteristic of the spreading of point perturbations.

Further computational saving may be made at the expense of resolution. If y is a low resolution representation of the model state (i.e. the dimension of y is less than that of x) defined as,

$$y = Yx$$

then with loss of resolution

$$x \doteqdot s + Y'(YY')^{-1} y = s + Xy$$

say, where s is some vector belonging to the null space of Y. Then $A_{\tau-1} P^{-1} X$ can be computed from

$$A_{\tau-1}(P^{-1}(x_\tau + X\delta_i) - P^{-1}(x_\tau))/\|\delta_i\|$$

and with a sacrifice of resolution

$$A_{\tau-1} P_\tau^{-1} \doteqdot A_{\tau-1} P_\tau^{-1} XY.$$

Thus the number of backward timesteps required depends on how much resolution of the error structure can be reasonably foregone. An obvious application of this would be to use a coarse grid for the error structure in a grid model. But a more significant possibility is the use of a transformation to avoid resolving gravity waves.

The influence of each constraint will spread with time. In a primitive-equation model most of this spread is associated with rapidly propagating gravity waves. This is only a nuisance since the waves carry no information if their amplitude is always negligible. Experimentally it is observed that

local errors spread only slowly, the main change usually being horizontal displacement (advection by the wind), and sometimes amplification by growth of disturbances. The rapid gravity wave spread could be taken care of with the above technique, by a transformation to vorticity $y = Yx$. Then gravity waves are not resolved since they are essentially irrotational. On the other hand the large-scale meteorology is retained. The perturbations δ_i are then point vortices whose representations $X\delta_i$ can be kept in a table rather than computed on demand. The sparse-matrix approximation will still require a procedure for trimming constraint domains to a manageable size, possibly by simply ignoring small effects outside a suitable radius. Since the main feature of the dynamics is advection of vorticity, spreading of a vorticity perturbation is a relatively slow process, so that constraints which are initially localized will remain localized for a significantly long time.

Each new item of data makes a contribution to the precision, represented as a new constraint. The effect of stochastic forcing will be that the informative value of any such constraint will decline toward zero. The demands of economy will dictate a retirement age; old constraints could be discarded when their contribution to the precision is no longer significant.

3.4. *Stochastic forcing*

The prime physical mechanism for random forcing is the interaction of the flow with processes not resolved by the model. It will be supposed that the resulting uncertainty can be represented by the generalization of eq. (11),

$$x_\tau = P(x_{\tau-1} + e_{\tau-1})$$

where the e_τ are random variables with zero means, uncorrelated with x_τ for all τ. For this it is sufficient that the mean of e_τ conditional on x_τ be zero. The covariance of e_τ conditional on x_τ is here allowed to be a function of x_τ. Pitcher (1977) considered similar forcing. To the author's knowledge, little is known about the detailed form that such forcing should take, but, if the implied growth of uncertainty is not to be unrealistically small, it should at least excite all the available instabilities. Hence the proposition;

Hypothesis (1): Stochastic forcing is of full rank, i.e. there is no state which is orthogonal to the forcing (in the low-resolution subspace, at least). This implies a strict limit on the norm of precision, and strict attenuation of precision by the forcing as follows. If e_τ has covariance S_τ then $x_\tau + e_\tau$ has precision

$$\hat{E}_\tau = (E_\tau^{-1} + S_\tau)^{-1}$$

but for any x

$$x'(E_\tau^{-1} + S_\tau)x \geqslant \lambda\|x\|^2$$

where λ is the smallest eigenvalue of S_τ, and λ is greater than zero as a consequence of hypothesis (1), so that $||\hat{E}_\tau|| \leqslant 1/\lambda$.

If the timesteps are short, S will be small and the following approximation will be computationally convenient,

$$\hat{E} \doteqdot N'EN$$

where

$$N = I - \tfrac{1}{2}SE$$
$$= I - \tfrac{1}{2}SA'A$$

and

$$\|N\| < 1$$

so that N represents an attenuation of precision, and computation of the evolution of covariance takes the final form

$$A_\tau = \begin{bmatrix} A_{\tau-1}N_{\tau-1}P_\tau^{-\frac{1}{2}} \\ R_\tau D_\tau \end{bmatrix}. \tag{13}$$

From eq. (13) the precision at time τ can be written as a sum:

$$E_\tau = \sum_{\nu=1}^{\tau} M_\nu' \Delta E_\nu M_\nu + E_0$$

where $\Delta E_\nu = D_\nu' R_\nu' R_\nu D_\nu$ is the increment of precision at time ν, and $M_{\nu-1} = \prod_{\mu=\nu}^{\tau} N_{\mu-1}P_\mu^{-1}$ is the cumulative effect of stochastic-dynamics on the precision. E_0 is the fixed component of precision, which is primarily that due to dynamical constraints (see Section 3.5). Now it is universal experience that the predictive value of data quickly declines as its age increases, so that there is little loss if old data is forgotten, but this implies

$$E_\tau \doteqdot \sum_{\nu=\tau-\kappa}^{\tau} M_\nu' \Delta E_\nu M_\nu + E_0$$

for some age κ. Note that this limits the radius of

precision, for if the radius increases by Δr in unit time then

$$\text{radius } (E_\tau) = \max\{\text{radius } (E_0), \text{radius } (\Delta E) + \kappa \Delta r\}$$

There is a considerable class of random forcing which will tend to localize the precision even in the absence of localized observations. For instance, there will be a localizing effect of this sort if S is diagonally dominant.

Errors introduced by deficiencies in the prediction model itself are not of the above kind. Imperfections of parameterization and numerical method are by definition 100% correlated with the model state. This is a type of bias which inevitably degrades accuracy, so in a simulation it would be appropriate for N to incorporate extra attenuation.

3.5. *Dynamical constraints*

Error in the winds cannot be independent of error in the temperatures, since geostrophic adjustment produces a balance which relates them to each other. To the author's knowledge, there is no good reason for supposing that fast-propagating gravity-wave disturbances do not exist in the atmosphere, but only that the amplitudes of these modes are small. This state of affairs is expressed in the spirit of the stochastic formulation by summarizing what is known about the variances of gravity-wave components in the form of a "weak" constraint on these modes, and then setting out to estimate the state of the atmosphere, including an estimation of the gravity-wave components, from the available data.

A "weak" constraint of this type was used by Sasaki (1970). This was a constraint on the tendencies of geopotential and wind, which can be expressed as follows. If T and t are partitions of A and a containing the rows corresponding to the dynamical constraint, i.e.

$$A = \begin{bmatrix} M \\ T \end{bmatrix} \quad \text{and} \quad a = \begin{bmatrix} m \\ t \end{bmatrix}$$

where M, m are merely the partitions containing the remainder of these arrays, so that

$$\|A(h-g) - a\|^2 = \|M(h-g) - m\|^2 + \|T(h-g) - t\|^2 .$$

then the Sasaki (1970a) constraint can be written

$$t = -Tg$$

$$Th = \begin{bmatrix} \eta_1 \, \partial\Phi/\partial\tau \\ \eta_2 \, \partial w/\partial\tau \end{bmatrix} \tag{14}$$

where η_1, η_2 are weight parameters, Φ the geopotential, w the horizontal component of wind, τ the time, and $\partial\Phi/\partial\tau$, $\partial w/\partial\tau$ are expressed by a prediction equation in terms of spatial gradients.

One disadvantage of this constraint is that it constrains the amplitudes of both the divergence and the vorticity of a wave, even though it is intended to control only the essentially irrotational gravity-wave modes. However, it is possible to devise constraints of the same type which are more appropriate and still correspond to localized precision. Indeed, the above constraint can be modified to affect directly only the divergence, as follows. We introduce an extra scalar field to be estimated, Ψ say, and instead of eq. (14) we set

$$Th = \begin{bmatrix} \eta_1 \, \partial\Phi/\partial\tau \\ \eta_2(\partial w/\partial\tau - k \times \nabla\Psi) \end{bmatrix} \tag{15}$$

where k is the unit vector normal to the horizontal. Since Ψ appears in the partition T, but nowhere else, it follows that the solution $\hat{\Psi}$, which minimizes ξ in eq. (4), is the stream function of the wind tendency, so that $\partial w/\partial\tau - k \times \nabla\hat{\Psi}$ is the irrotational component of the wind tendency. The degree of constraint on the amplitude of divergence of a wave is then a function of the phase speed of the wave.

More selective constraints which suppress the fast-propagating modes to a greater degree can be specified by substituting higher order time derivatives in place of the first-order derivatives in eq. (14) or (15). Whatever order time derivatives are specified, these equations correspond to an hypothesis that the components of Th have unit variance and zero mean, so that their magnitudes are probably of order unity. This hypothesis can be justified on the ground that it is a direct application of the Kreiss (1980) "bounded derivative" principle; that slowly varying solutions of the dynamical equations are possible if and only if time derivatives are bounded. Alternative ways of applying the bounded derivative principle, using scale analysis are discussed by Browning et al. (1980), and "weak" constraints could also be formulated in those ways.

3.6. *Boundary constraints*

It will usually be desirable to estimate variables at any boundaries of the analysed region, so that the estimate will adjust to data near the boundaries. The simplest constraint merely ties the estimate close to the "first guess" at the boundary, but experience shows that an additional constraint on the gradient of wind normal to the boundary is advisable to avoid "noise" if the boundary is to be allowed to change much, especially in the tropics. If B and b are partitions of A and a containing the boundary constraints in the same way that A is partitioned in eq. (14), then the constraint which was found to be satisfactory is

$$B(h-g)-b=\begin{bmatrix}\beta(h-g)\\ \gamma n\cdot\nabla u\\ \gamma n\cdot\nabla v\end{bmatrix} \tag{16}$$

where β, γ are parameters, u, v are the two components of the wind w, and $n\cdot\nabla$ is the gradient in the direction normal to the boundary.

4. Test for rejection of bad data

Any practical scheme must include some means of rejecting, as far as possible, data which is corrupted by errors of communication etc. If the assumptions made for eq. (2) are valid, and if the errors e in eq. (2) have a normal distribution, then the residual sum of squares ξ (in eq. (3) or (4)) has a chi-squared distribution with $n-m$ degrees of freedom, where n is the total number of constraints, and m is the number of variables being estimated. This gives a reasonable test of the hypothesis that the data comes from a distribution of the kind implied by the model: data should be rejected only if this chi-squared statistic is larger than an appropriate significance level threshold. This test leaves unanswered the question of which items to reject, in the event that rejection is indicated. A simple though approximate procedure was used in the experiments as follows. First, an analysis was done using all the data, but to economize, only a low precision was specified for the solution. Second, the difference between each observation and its analysis estimate was computed, and the observation was rejected if this difference was larger in magnitude than a test constant times its observational error standard deviation. Third, the final analysis was computed without the rejected data. The test constant was determined by the desired size of the test: the probability of rejecting good data.

The true size of this test is less than the nominal size, for if an analysis without using one item yields an estimate of it of d_0 with variance v_0, then an analysis using the item d_1, sampled from an independent distribution with variance v_1, will yield an estimate d_2 where

$$d_2=(v_0\,d_1+v_1\,d_0)/(v_0+v_1)$$

and the variance of the difference d_2-d_1 is

$$v_1(v_1/(v_0+v_1))<v_1.$$

It is here assumed that the estimates d_0, d_1, d_2 are unbiased.

5. Computational method of minimization

The method used for the solution of eq. (4) is an iteration which solves for all variables simultaneously. For the experimental analyses these are values of geopotential and two wind components at all nodes of a grid. A type of block relaxation is used since simpler schemes appear to converge too slowly. All variables in a spatially localized patch are relaxed simultaneously, by direct solution of the linear equations while holding surrounding values fixed. It has been found that for the two-dimensional problem an adequate convergence rate is achieved using only the four mesh points surrounding a single cell of the grid in each patch. In three dimensions the vertical column of such points could be relaxed simultaneously. For the solution of $Eh=r$ the iteration can be expressed as

$$h_{n+1}=h_n+\lambda\,(r-(E-L_n)h_n-L_n\,h_{n+1}) \tag{17}$$

where M_n, L_n vary cyclicly with n,

$$M_n=E-L_n-L_n'$$

and for each n there is a row and column permutation K for which KL_nK^{-1} is lower-triangular and KM_nK^{-1} is block-diagonal, the block $M_{n,k}$ corresponding to the kth patch to be relaxed in the nth pass.

Schemes of this type are discussed at length in Wachspress (1966). The main point to note is that convergence of the iteration is assured for the relaxation parameter range $0<\lambda<2$ if E is

positive definite. In eq. (5) $E = D'R'RD + A'A$ and the number of constraints (number of rows of A) exceeds the dimension of h, so it is sufficient that A have full rank. In the experiments it is thus sufficient that the three sets of constraints on vorticity, divergence, and a geostrophic geopotential gradient, span the domain space of A. In practice no special precautions seem to be necessary for the linear problem, except for a careful choice of finite-difference operators as noted below. The convergence rate has been found to be slow but satisfactory in practice, a maximum 300 mb wind correction error was of the order of 1.5 m/s after six complete cycles, each cycle consisting of four passes in a chequerboard pattern. Wachspress gives a valuable discussion of the factors affecting convergence rate. In the present circumstances a suitable value for λ is about 1.7. To solve eq. (5), the relaxation of the kth block is

$$h_{n+1,k} = h_{n,k} + \lambda M_{n,k}^{-1}(A'A(g - h^*) + D'R'R(d - Dh^*) + A'a)_k \qquad (18)$$

where $h^* = h_n$ for the first block, but subsequently $(h^*)_k$ is overwritten by $h_{n+1,k}$ as these are computed.

The choice of finite difference approximations for the constraints is critical. It is important that there is no mesh redundancy as described by Platzman (1958). This is because "noise" which the finite difference operators do not "see" will not be removed in the solution. For this reason the operators chosen for the experiments correspond to the staggered mesh shown in Fig. 1. The usual centred difference approximations for gradients were used. For each data item the interpolation operator used references the four nearest grid values of the same type, to implement bilinear interpolation.

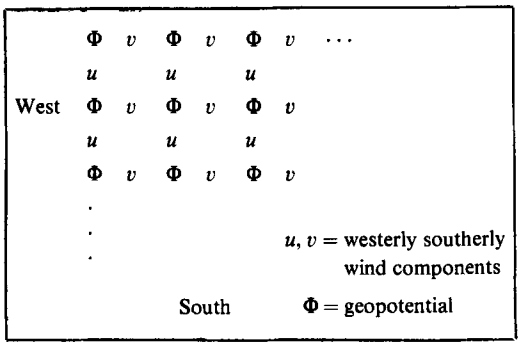

Fig. 1. Computational grid.

Non-linearities do not seem to introduce unreasonable difficulties. Probably the most natural way of generalizing the method is to use linearized approximations of the constraints, and compute an outer iteration to adjust the non-linear effects. This outer iteration is known as the Gauss, or Gauss–Newton iteration for the solution of a non-linear least-squares problem. For the minimization of $\|Ah - r\|^2$ the outer iteration can be written

$$G_m'(G_m(h_{m+1} - h_m) + Ah_m - r) = 0 \qquad (19)$$

where the rows of G_m are linearized versions of the constraints, linearized about h_m. In the special case where A is linear, $G_m = A$. G_m is known as the (Frechet) derivative of A at h_m. Conditions which would guarantee the convergence of this iteration, and be sufficiently sharp to be useful in practice, are not known. It is clear on general grounds that a unique minimum may not exist if the model is too strongly non-linear, or if the minimum value of $\|Ah - r\|^2$ is too large. It may be useful to note that since a linear solution will be a fairly good approximation of the final solution, it would not be unreasonable to apply the chi-squared test for rejection of bad data after every cycle of the Gauss–Newton iteration, which would guarantee a limit on the magnitude of the residual. The model envisaged is only weakly non-linear, for on the meteorological scale envisaged it is usual for non-linearities in the dynamical constraints to be in the nature of a significant but relatively small correction to a linear solution, and the observation operators for the usual data types are linear, or nearly linear. In the experiments a spatial filter was applied to the non-linearities to facilitate computation, with apparently good results.

6. The experiments

6.1. *Aims*

The experiments were restricted to a simple model in which precision is stationary in time and in space. They were intended first, as a trial of the computational method of minimization, and second, to investigate a simple, stationary, localized model of precision, since if localized precision is

more apt than localized covariance in the general case, it might also be more apt for a stationary analysis scheme. This is intended only as preliminary exploration in unfamiliar territory.

6.2. *A stationary model of precision for the experiments*

In view of the considerations of Section 3.5, the experimental precision will be formulated for two models, 1a and 1b, in terms of vorticity error, gradients of vorticity error, and the balance constraint of eq. (14). The constraint of vorticity difference is contained in a partition Q of A, where, for model 1a,

$$Q(h-g) = \kappa_1(\zeta(h) - \zeta(g)). \tag{20}$$

This forces the estimated vorticity close to the guess vorticity, where κ_1 is a weight parameter and ζ is the vertical component of vorticity $\nabla \times w$. For model 1b,

$$Q(h-g) = \begin{bmatrix} \kappa_1(\zeta(h) - \zeta(g)) \\ \kappa_2 \nabla(\zeta(h) - \zeta(g)) \end{bmatrix} \tag{21}$$

where κ_2 is an additional parameter to control the gradient of vorticity.

The tendency equations substituted into eq. (14) were, for the tendency of geopotential,

$$\partial\Phi/\partial\tau = -\Phi\nabla \cdot w \tag{22}$$

where ∇ is the gradient operator, and for the tendency of the horizontal component of wind,

$$\partial w/\partial\tau = (f + \overline{\nabla \times w}) \times w - \nabla(\Phi + \tfrac{1}{2}\overline{w} \cdot \overline{w}) \tag{23}$$

where f is the coriolis vector, and an overbar denotes filtering of terms which were held constant in the inner iteration. This treatment of nonlinearities is not exactly as in eq. (19), because of limitations of memory address-space in the small computer used. Instead, it is an approximation based on the observation that eq. (23) is analogous to the geostrophic relation with a correction to the coriolis parameter and a correction to the geopotential. In practice convergence of the outer iteration was rapid, only a few cycles being necessary, and the inner iteration required less computation on each successive cycle. The net requirement was about 80% more computation than for a linear solution. As an additional control the term $\overline{\nabla \times w}$ was limited to the range $\{2/3f, 4/3f\}$. This may have been excessively cautious.

6.3. *Method*

Real data and guess fields provided by an operational prediction model on the area 130 E to 160 W and 25 S to 55 S were used on a 2.5 degree latitude–longitude grid. Although the results are particular to the distribution of error of this model, it is thought that this is likely to be representative enough of numerical models to be interesting. Parameters for models 1a and 1b were chosen empirically so as to fit these data as well as possible. The details of how this was done are thought not to be sufficiently relevant to warrant inclusion here, but the results are listed in Table 1. These parameters were substituted into models 1a, 1b, for the second part of the experiment.

To compute some representative covariance patterns implied by model 1a, analyses of single artificial "observations" of height and wind located at the centre of the grid were computed, the precision of these artificial "observations" being small, and the guess consisting of zero wind and a constant geopotential. From eq. (6) it is evident that if $D'R'RD$ is small relative to $A'A$, and if RD picks out the ith component of g, i.e. $RDg = \varepsilon g_i$, for small ε, then the solution $h - g$ is proportional to the covariance of the guess error with g_i. The resulting patterns are shown in Fig. 2.

A comparison was made between the models 1a, 1b, and a model which is known to perform well, as follows. All schemes were tested on the same sample of data and guess fields. A random selection of the wind observations was withheld in each case; the probability of an observation being withheld was 15%. Then the analyses were

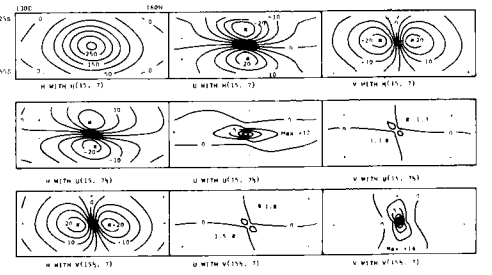

Fig. 2. Covariance of error in the guess for model 1a (Table 1). Units for height, wind are m, m s^{-1}. Because of the grid staggering the centre is (15, 7) for h, (15, 7½) for u and (15½, 7) for v.

computed without using the withheld observations, and the withheld winds compared with the estimates.

The comparison scheme (model 0) consists of Cressman (1959) analyses of geopotential, southerly and westerly wind components separately. The three fields are then adjusted to a mutual balance by minimization of the functional

$$\xi = \sum_i (\alpha_1 S_{1,i}(\Phi_i - \hat{\Phi}i)^2 + \alpha_2 S_{2,i}(w_i - \hat{w}_i)^2 + \beta(\partial w_i/\partial\tau)^2) \quad (24)$$

the summation being over indexes of all grid points, where $\partial w/\partial\tau$ is given by eq. (23), w is the analysis wind, $\hat{w}$ the Cressman analysis wind, Φ is the geopotential, S_1, S_2 are sums of Cressman weights over all scans, having spatial variation depending on the data density, α_1, α_2, β are parameters for relative weight on geopotential, wind, and the coupling between geopotential and wind, and $w_i^2 = u_i^2 + v_i^2$, for wind components u, v. The solution of eq. (24) is computed by the same method as model 1, except that the relaxation blocks contain only three components, one from each of Φ, u, v at each grid point. Convergence is fast.

6.4. *Results*

The covariance patterns shown in Fig. 2 are not isotropic, and they are not nearly as localized as the precision, which has a radius of support of $2^{1/2}$ grid intervals. Nevertheless, the covariance between very distant points is very small, as is assumed in many analysis schemes.

The differences between models were small. The root-mean-square wind errors for a sample of 181 withheld wind observations, accumulated over about 50 analysis cases, were as follows: 11.018 m/s for model 0, 11.267 m/s for model 1a, 11.273 m/s for model 1b.

In model 1a the error distribution can be described as follows: From Table 1 the variance of circulation (the line integral of tangential wind) around an area Δ is $\Delta/25{,}000$ (S.I. units).

The results for the error variance of wind are contradictory. From Fig. 2 the standard deviation of error in the guess wind (at the centre of the grid) is only 5.1 m/s. The error in an analysis by model 1a should therefore not exceed this. A total error of 11.267 m/s has been observed so that a standard deviation of error in the observational measurements of at least 10 m/s is implied, but only 7.1 m/s has been assumed (Table 1). The "chi-squared" statistic for a discrepancy of this magnitude from 181 independent samples is significant at the 0.01% level; the difference is too great to attribute to sampling variation. It is therefore possible in principle to estimate more of the parameters of the model: Since the observational error variances are not well known they could be diagnosed so as to fit. Such an exercise might be of value for selection of a model from a set of tentative ones.

Table 1. *Model parameters (S.I. units)*

A priori parameters, models 1a and 1b

Observation type	Rejection criterion	Standard deviation σ of observation error (includes interpolation error)
Wind	3.5σ	7.1 m s^{-1}
Height of 300 mb surface	3.5σ	22 m

Optimized parameters

Model 0		Model 1a		Model 1b	
α_1	0.0025	η_1	5.0	η_1	5.0
α_2	0.1	η_2	6.5×10^3	η_2	6.5×10^3
β	1.35×10^4	κ_1	2.5×10^4	κ_1	2.5×10^4
				κ_2	2.0×10^8

7. Conclusions

Approximate localization of covariance of error in an estimate of the state of the atmosphere cannot be guaranteed with typical observational networks, particularly in data-sparse areas, even though localization is essential for computational economy. Instead, approximate localization of precision is to be expected in reasonable circumstances. Then the conventional covariance representation of the distribution of error is probably not the most appropriate. It has been noted that analysis of independent data items corresponds to addition of precision. Consequently, the precision of the least-squares estimate remains localized if the precisions of the inputs are localized. A prediction model will spread both the precision and the covariance. But the precision is attenuated by the growth of uncertainty, so that young independent localized observations can dominate, maintaining approximate localization of precision.

A major advantage of carrying precision rather than covariance is that data assimilation corresponding to addition of precision is a simple operation. Another advantage is that there is no need to analyse wind and mass fields separately, since the coupling between them can be specified in terms of localized precision.

Even as a stationary analysis scheme, one merit of the formulation is that it naturally allows direct analysis of any data type. For example, in a three-dimensional analysis it would be both unnecessary and undesirable in principle to first estimate temperatures from satellite-measured radiances as is the current practice. Instead, we define an observation operator, which gives expected radiance as a function of the model state. The analysis then adjusts this state so as to fit the observed radiances, and all the other data.

The experiments demonstrate that a simple analysis model formulated in terms of precision is a practical possibility, and show some of its features.

8. Appendix

Evaluation of covariance in the counter-example

Using the notation established in the example of Section 3.1, define a change of basis:

$$\Phi_j = \sum_{k=0}^{n-1} T_{jk}\,\hat{\Phi}_k, \quad j = 0, 1 \ldots, n-1$$

$$T_{jk} = (\exp(2\pi ijk/n))/n^{1/2}, \quad i = (-1)^{1/2}$$

Let T^* be the complex adjoint of $T = [T_{jk}]$, and note that $T^* = T^{-1}$. Define $\hat{Q} = T^* Q T$, and note that $\hat{Q}$ is diagonal, with diagonal elements

$$\hat{Q}_{jj} = 1 + 2\alpha\cos(2\pi j/n) = \lambda_j$$

and

$$Q = T\hat{Q}T^*.$$

The observation operator for an observation at $j + \frac{1}{2}$ is D_j,

$$D_{jk} = \begin{cases} 1 & \text{if } k = (j+1) \bmod (n) \\ -1 & \text{if } k = j \\ 0 & \text{otherwise} \end{cases}$$

so that

$$D_j = \hat{D}_j T^*$$

where

$$\hat{D}_{jk} = (\exp(2\pi ik(j+1)/n) - \exp(2\pi ikj/n))/n^{1/2}$$

and

$$(1/\sigma^2)D^*D = \sum_{j=0}^{n-1} (1/\sigma^2)D_j^* D_j$$

$$= T(1/\sigma^2)\hat{D}^*\hat{D}T^*.$$

Note that $\hat{D}^*\hat{D}$ is diagonal, with diagonal elements

$$(\hat{D}^*\hat{D})_{jj} = 2(1 - \cos(2\pi j/n)) = 2\mu_j.$$

Then

$$S = (\hat{Q}^{-1} + (1/\sigma^2)\hat{D}^*\hat{D})^{-1}$$

is diagonal, with diagonal elements

$$S_{jj} = \sigma^2\lambda_j/(\sigma^2 + 2\lambda_j\mu_j)$$

and

$$((Q^{-1} + (1/\sigma^2)D'D)^{-1})_{m,k} = (TST^*)_{m,k}$$

$$= (1/n)\sum_{j=0}^{n-1} \exp(2\pi ij(k-m)/n)S_{jj}.$$

Note that $\lambda_0 = 1 + 2\alpha$ and $\mu_0 = 0$, so that the result is as stated in Section 3.1.

REFERENCES

Browning, G., Kasahara, A. and Kreiss, H. O. 1980. Initialization of the primitive equations by the bounded derivative method. *J. Atmos. Sci. 37*, 1424–1436.

Cressman, G. P. 1959. An operational objective analysis scheme. *Mon. Wea. Rev. 87*, 367–374.

Epstein, E. S. and Pitcher, E. J. 1972. Stochastic analysis of meteorological fields. *J. Atmos. Sci. 29*, 244–257.

Gandin, L. S. 1963. *Objective analysis of meteorological fields.* Translated from Russian (1965) by Israel Program for Scientific Translations, Jerusalem, 242 pp.

Kreiss, H. O. 1980. Problems with different time scales for partial differential equations. *Comm. Pure Appl. Math. 33*, 399–439.

Laurmann, J. A. 1978. A small perturbation approximation for stochastic dynamic weather prediction. *Tellus 30*, 404–417.

Leith, C. E. 1974. Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev. 106*, 409–418.

Pitcher, E. J. 1977. Application of stochastic dynamic prediction to real data. *J. Atmos. Sci. 31*, 3–21.

Platzman, G. W. 1958. The lattice structure of the finite-difference primitive and vorticity equations. *Mon. Wea. Rev. 86*, 285–292.

Sasaki, Y. 1970a. Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev. 98*, 875–883.

Sasaki, Y. 1970b. Numerical variational analysis formulated under the constraints as determined by longwave equations and a low-pass filter. *Mon. Wea. Rev. 98*, 884–898.

Sasaki, Y. 1970c. Numerical variational analysis with weak constraint and application to surface analysis of severe storm gust. *Mon. Wea. Rev. 98*, 899–910.

Wachspress, E. L. 1966. *Iterative solution of elliptic systems.* Engelwood Cliffs, N.J.: Prentice-Hall.

Wahba, G. and Wendelberger, J. 1980. Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon. Wea. Rev. 108*, 1122–1143.