INTRODUCTION TO NUMERICAL WEATHER PREDICTION DATA ASSIMILATION METHODS

P Courtier

1. THE PROBLEM

In numerical weather prediction data assimilation consists of the process which estimates the initial conditions of the forecast using all the available information. A description of the current observing system can be found in *McGrath* (1993); around 10^5 elementary pieces of information are currently used over 24 hours by the operational data assimilation system.

The current operational ECMWF model (*Simmons*, 1991) has a horizontal resolution of 90 km and covers the whole globe. On the vertical the atmosphere is sampled with 31 levels from the surface up to 10 hPa. The number of degrees of freedom of the model is then of the order of 10^7.

Over 24 hours, the estimation problem is clearly underdetermined. The time dimension is thus a critical element of any data assimilation system: it is essential to carry forward in time information from past observations using the forecast model since it is the best information propagator available.

We have identified the main difficulties of data assimilation:

- large dimension problem
- time dimension critical but non linear dynamic
- observations of variable nature and quality

The purpose of this paper is to provide the theoretical basis of the algorithms used in operational meteorology together with some recent developments. Most of the material presented here is well documented in the meteorological literature (*Lorenc*, 1986; *Ghil and Manalotte-Rizzoli*, 1991; *Daley*, 1991) or in other fields under the generic name of inverse problems (*Tarantola*, 1987).

In the next section we shall introduce linear estimation theory. Section 3 briefly describes the Optimal Interpolation algorithm and then concentrates on the variational approach. Section 4 deals with the time dimension; it theoretically introduces the Kalman filter then the four-dimensional variational assimilation algorithm is described, together with some simplifications, which allow an operational implementation in the foreseeable future. Finally, an algorithm is presented as an attempt to provide error bars from the variational algorithms and to improve the treatment of the time dimension.

2. INTRODUCTION TO LINEAR ESTIMATION

In the following the random variables are underlined whereas realizations of these random variables are not.

2.1 A simple illustration

Let us assume that we are interested in an estimation of the temperature of the ECMWF Lecture Theatre. We consider the room temperature as a random variable T of expectation $\langle T \rangle$ and variance $\langle (T-\langle T \rangle)^2 \rangle$ which are not supposed to be known.

We have a thermometer of known precision σ_o and we read T_o. The measurement T_o is considered as one realization of a random variable T_o of expectation $\langle T \rangle$ (the observation is assumed unbiased) and variance $\langle (T_o-\langle T \rangle)^2 \rangle = \sigma_o^2$.

In the absence of any other information, the Best Linear Unbiased Estimator (BLUE) of the room temperature is $T_a = T_o$ and for the particular realization T_o of the measurement we have $T_a = T_o$.

In other words, we have been able to produce a rather pedantic presentation of what you should do when you want to have an idea of the room temperature - simply to read the thermometer; at least it is not against common sense!

However, if you look carefully at the ECMWF attendees of a presentation in the Lecture Theatre, most of them are wearing a jumper: they have a priori (background) information T_b on the room temperature (it's gonna be cold!). T_b is considered as one realization of a random variable T_b of expectation $\langle T \rangle$ and variance $\langle (T_b-\langle T \rangle)^2 \rangle = \sigma_b^2$.

Intuitively, one realizes that the background information, combined with the observation, should lead to a better estimate than the observation alone. We then look for the BLUE, a linear estimate

$$T_a = \alpha T_b + \beta T_o \tag{1}$$

which is unbiased and best (of minimum error variance). As the background information and the observations are assumed unbiased, we get

$$\langle T \rangle = \alpha \langle T \rangle + \beta \langle T \rangle$$

and then $\alpha + \beta = 1$.

In meterological practice (1) is usually rewritten as

$$T_a = T_b + \beta (T_o - T_b).$$

We look for a correction to the background T_b which is a linear function of the difference observation minus background.

β is evaluated so as to fulfil the minimum variance criterion for the estimate

$$\begin{aligned} <(T_a-<T>)^2> &= (1-\beta)^2<(T_b-<T>)^2> + \beta^2<(T_o-<T>)^2> \\ &=(1-\beta)^2\ \sigma_b^2 + \beta^2\ \sigma_o^2 \end{aligned} \qquad (2)$$

which is minimum for $\beta = \dfrac{\sigma_b^2}{\sigma_b^2+\sigma_o^2}$

and $$T_a = T_b + \frac{\sigma_b^2}{\sigma_b^2+\sigma_o^2}(T_o-T_b)$$

and for the actual realizations T_o and T_b,

$$T_a = T_b + \frac{\sigma_b^2}{\sigma_b^2+\sigma_o^2}(T_o-T_b)$$

both pieces of information are weighted according to their statistical quality. If we consider the limit case of a very low quality measurement ($\sigma_o>\sigma_b$), then the analysis remains equal to the background. On the other hand, if the observation is of very high quality ($\sigma_o<\sigma_b$), the analysis is equal to the observation.

The variance of analysis error is found by replacing β with its expression in (2)

$$<(T_a-<T>)^2> = \left(\frac{1}{\sigma_o^2}+\frac{1}{\sigma_b^2}\right)^{-1}.$$

2.2 General formulation of linear estimation

The state of the model considered as a random variable is denoted by $\underline{x}$. As already said, for real size problems, it is a vector of length 10^7. Observations $\underline{y}$ are available under the form

$$\underline{y} = H\underline{x} + \underline{\epsilon}. \qquad (3)$$

H is the so-called observation operator and $\underline{\epsilon}$ the observation errors (representativeness and instrumental as explained by *Lorenc*, 1986). The observations are assumed unbiased ($<\underline{\epsilon}>=0$) and of known error covariances $<\underline{\epsilon}\ \underline{\epsilon}^t> = O$.

A background $\underline{x}_b$ is available unbiased $\langle \underline{x}_b \rangle = \langle \underline{x} \rangle$ and of known error covariance

$$\langle (\underline{x}_b - \langle \underline{x} \rangle)\ (\underline{x}_b - \langle \underline{x} \rangle)^t \rangle = \boldsymbol{B}.$$

The Best Linear Unbiased Estimator is then given by

$$\underline{x}_a = \underline{x}_b + \boldsymbol{K}(\underline{y} - \boldsymbol{H}\underline{x}_b) \tag{4}$$

with $$\boldsymbol{K} = \boldsymbol{B}\,\boldsymbol{H}^t(\boldsymbol{H}\,\boldsymbol{B}\,\boldsymbol{H}^t + \boldsymbol{O})^{-1} = (\boldsymbol{B}^{-1} + \boldsymbol{H}^t\boldsymbol{O}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{O}^{-1}. \tag{5}$$

The covariances of analysis error

$$\langle (\underline{x}_a - \langle \underline{x} \rangle)(\underline{x}_a - \langle \underline{x} \rangle)^t \rangle = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{B} = (\boldsymbol{B}^{-1} + \boldsymbol{H}^t\boldsymbol{O}^{-1}\boldsymbol{H})^{-1}. \tag{6}$$

This classical result of linear estimation is clearly established in several ways in *Jaszwinski* (1970) or in *Lorenc* (1986) and by various other authors. A more general formulation where a background is not necessarily included is presented (in a very accessible way) in *Tarantola* (1987). The analysis value is obtained for a given realization of the background and the observations replacing the random variables in (4) by their values.

In Meteorology the practical difficulty is that it is impossible to use (4) and (5) directly. $\boldsymbol{B}$, for example, is a matrix of size 10^7 x 10^7 which is about 1000 times the total archiving capabilities of ECMWF and one million times the memory size of the current computers.

The scientific difficulty of data assimilation is then to find algorithms which simplify (4) and (5) to an affordable amount of computer resources, while preserving some of the essential characters.

2.3 Quasi linear case

The previous results can easily be generalised when the observation operators are weakly non linear: the tangent linear approximation $\boldsymbol{H}'$ is assumed to be valid for the order of magnitude of the background errors:

$$\begin{aligned} \underline{y} &= \boldsymbol{H}(\underline{x}) + \underline{\varepsilon} \\ &= \boldsymbol{H}(\underline{x}_b) + \boldsymbol{H}(\underline{x}) - \boldsymbol{H}(\underline{x}_b) + \underline{\varepsilon} \\ &= \boldsymbol{H}(\underline{x}_b) + \boldsymbol{H}' \cdot (\underline{x} - \underline{x}_b) + \underline{\varepsilon}. \end{aligned}$$

Equations (4), (5) and (6) then become

$$\underline{x}_a = \underline{x}_b + \boldsymbol{K}(\underline{y} - \boldsymbol{H}(\underline{x}_b)) \tag{7}$$

with $$\boldsymbol{K} = \boldsymbol{B}\,\boldsymbol{H}'^t(\boldsymbol{H}'\boldsymbol{B}\boldsymbol{H}'^t + \boldsymbol{O})^{-1} \tag{8}$$

and $$\langle (\underline{x}_a - \langle \underline{x} \rangle)(\underline{x}_a - \langle \underline{x} \rangle)^t \rangle = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H}')\boldsymbol{B}. \tag{9}$$

3. TWO PRACTICAL IMPLEMENTATIONS OF LINEAR ESTIMATION

3.1 Optimal Interpolation (OI)

OI was introduced in Meteorology by *Gandin* (1963) and significantly contributed to the progress of numerical weather forecast quality during the late 70's and the 80's. The first ECMWF implementation is described in *Lorenc* (1981) and a revised implementation in *Shaw et al* (1987).

The basic idea to simplify (4) and (5) is that, for a given geographical location, only the neighbouring observations are useful. This argument relies on the relatively small horizontal length scale (500 km) of the height forecast error correlation (see *Hollingsworth and Lönnberg*, 1986 and *Lönnberg and Hollingsworth*, 1986).

For each location a small set of predictors is kept and the same equation as (4) and (5) is solved but with matrices of tractable size. In some operational implementations wind and mass are analysed at a given horizontal location and at a given level with as few as 15 predictors. At ECMWF, a box of a typical size 500 km is considered and an order of magnitude of 700 predictors is kept.

The second approximation comes from the necessity of an analytical model for the forecast error covariances which have to be computed at the observation locations. This model in most (all?) implementations relies on isotropy and geostrophy on the f-plane (f is assumed instant locally) with a separability vertical/horizontal (the spatial correlations can be expressed as a product of two functions, one which depends on the horizontal distance only and one on the vertical distance only).

Generally speaking, the more local, the more noisy the analyses are (there are jumps in the data used from one location to another). A second weakness of OI is that if, in principle, any linear (or linearised) observation operator may be used, it is difficult to practically make use of observations indirectly related to the model parameters since in (5) all the forecast error covariances in observation space have to be computed. *Durand* (1985) implemented a direct use of satellite radiances but it significantly increased the PERIDOT OI code complexity.

The analysis error variances are easily computed using (6) for each simplified problem.

3.2 Variational formulation of linear estimation - 3D-Var

Coming back to the simple example of the room temperature estimation, if one introduces the function

$$J(T) = \tfrac{1}{2}\frac{(T-T_b)^2}{\sigma_b^2} + \tfrac{1}{2}\frac{(T-T_o)^2}{\sigma_o^2}$$

this function is minimal for

$$T_a = \frac{\sigma_o^2}{\sigma_o^2+\sigma_b^2} T_b + \frac{\sigma_b^2}{\sigma_o^2+\sigma_b^2} T_o$$

which is the same as the BLUE. Furthermore the second derivative of J is

$$J'' = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}$$

and is equal to the inverse of the covariance matrix of estimation error.

It is thus possible to find the estimate of the statistical problems minimizing a deterministic function which measures the misfit between the estimate and the information weighted by its statistical quality.

In other words, a linear regression and a least square fit provide the same answer. In the general case, let us consider the cost function

$$J(x) = \tfrac{1}{2}(x-x_b)^t B^{-1}(x-x_b) + \tfrac{1}{2}(Hx-y)^t O^{-1}(Hx-y) = J_b + J_o. \qquad (10)$$

Using the adjoint technique one can compute the gradient of the cost function and thus use an iterative minimization scheme suitable for large scale problems (quasi-Newton like M1QN3 from INRIA, *Gilbert and LeMarechal*, 1989). Here is the first approximation: for computational reasons, we shall allow 30 to 100 iterations (but not 1000!) and the convergence will not be achieved to machine accuracy. The OI scheme exactly solved a multitude of approximate problems. 3D-Var approximately solves the global problem.

However, one difficulty remains, namely the specification of the covariances of the forecast errors B: the spatial correlations are generally called "structure functions". As already said, dealing with the full matrix is intractable and one has to introduce simplifications. In both the ECMWF (*Courtier et al*, 1993) and the NMC (*Parrish and Derber*, 1992) implementations of J_b, the basic ideas are similar and already expressed in *Phillips* (1986). 3D isotropy and geostrophic coupling over the sphere using Hough mode separation (ECMWF) or a balance equation (NMC) are practically feasible while providing a theoretical improvement on top of the OI implementation. We now describe the key points of the ECMWF implementation while referring to *Courtier et al* (1993) for a detailed presentation.

3.3 Structure functions specification in 3D-Var

2D univariate homogeneous covariances

The basic property we are using is that, as homogeneous covariances over the sphere are by definition invariant with rotations, it becomes diagonal while expressed in the spherical harmonics basis (*Boer*, 1983). More precisely, the covariance between two points P and Q of the sphere is a function of only $\mu = \cos\theta$

where θ is the angle between P and Q. f may be expressed as a Legendre polynomial series

$$f(\mu) = \sum_n f_n P_n^o(\mu) \sqrt{2n+1}.$$

(The square root is here only for normalisation).

We then have the cost function expression J_b for a field $x(\lambda,\mu) = \sum_{n=0}^{N} \sum_{m=-n}^{n} x_n^m Y_n^m(\lambda,\mu)$ and a background field

$$x_b(\lambda,\mu) = \sum_{n=0}^{N} \sum_{m=-n}^{n} x_{bn}^m Y_n^m(\lambda,\mu)$$

$$J_b(x) = \sum_{n=0}^{N} f_n^{-1} \sum_{m=-n}^{n} |x_n^m - x_{bn}^m|^2$$

2D univariate, homogeneous correlations

Knowing the field $\sigma_b(\lambda,\mu)$, of the standard deviations of error, assuming the correlations homogeneous means that the covariances of $\frac{x(\lambda,\mu)}{\sigma_b(\lambda,\mu)}$ are homogeneous, we are back to the previous case f being a correlation verifies $f(0) = 1$ and then $\sum_{n=0}^{N} f_n(2n+1) = 1$.

The algorithm to compute J_b then becomes

i) from the spectral component x_n^m and x_{bn}^m of the control variable and the background respectively, computes the difference $\delta x_n^m = x_n^m - x_{bn}^m$

ii) transforms δx_n^m to grid point space and obtains the field $\delta x(\lambda,\mu)$

iii) divide $\delta x(\lambda,\mu)$ by the standard deviations $\sigma_b(\lambda,\mu)$

iv) transforms back to spectral space and obtains the spectral coefficients $\left(\frac{\delta x}{\sigma_b}\right)_n^m$

v) computes the cost function as $J_b(x) = \sum_{n=0}^{N} f_n^{-1} \sum_{m=-n}^{n} \left(\frac{\delta x}{\sigma_b}\right)_n^{m^2}$

3D univariate separable structure functions

As already said, assuming separable structure functions separable means that the correlation between two points can be expressed as a product of a function of the horizontal distance only and a function of the vertical distance only.

In 3D-Var, one may consider the vertical distance as a function of the hybrid vertical coordinate. After projection on the eigenvectors of the vertical correlation matrix, we are back to the 2D case. In optimal interpolation, the separability is assumed with pressure as vertical coordinate; this is different from 3D-Var where the vertical coordinate is terrain-following close to the surface. This leads to differences in the analysis increments near orography.

3D univariate, non separable

The idea is to have the vertical correlation matrix dependent on the total wavenumber n. More precisely, having obtained $\left(\frac{\delta x}{\sigma_b}\right)_n^m$ as in step iv of the 2D univariate case for all levels k (it is a vector of length k), one may consider a vertical covariance matrix $A(n)$ as a function of the total wavenumber n. The cost function then becomes

$$J_b(x) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \left(\frac{\delta x}{\sigma_b}\right)_n^{mt} A(n)^{-1} \left(\frac{\delta x}{\sigma_b}\right)_n^m$$

The implication of this formulation on the structure function in grid point space is discussed in *Phillips* (1986) and in the 3D-Var framework in *Courtier et al* (1993). This implementation of non separability can be seen as a kind of 3D isotropy with sharper vertical structures associated to small horizontal scales. The specification of the matrices $A(n)$ is achieved using statistics of the departure between a 24 h forecast and a 48 h forecast valid for the same time. The approach proposed by *Parrish and Derber* (1992) is described and validated by *Rabier and McNally* (1993).

Multivariate

$(x-x_b)$ is separated into a Rossby and a Gravity contribution. The latter is penalized ensuring 10% of the flow ageostrophic. For more details, we refer to *Courtier et al* (1993), in particular for some difficulties introduced by the vertical discretization while going back from geopotential to temperature and surface pressure. In the current implementation, it requires separability of the σ_b variation.

4. DEALING WITH THE TIME DIMENSION

We said in the introduction that the time dimension is a critical aspect for the specification of the initial condition of a numerical weather prediction. In this section we introduce an algorithm, the Kalman filter

(*Kalman*, 1960; *Ghil et al*, 1981) which provides a comprehensive and rigorous framework in the linear case. It may easily be extended to the quasilinear case and is then called extended Kalman filter (*Jaszwinski*, 1970).

4.1 The extended Kalman filter

Here we only present the algorithm; the theoretical results can be found in the above references and, in particular, in *Jazwinski* (1970). Denoting by $x(t)$ the state of the atmosphere at time t, we are able to propagate forward in time the information using the forecasting model M

$$x(t+T) = M(t+T,t)\, x(t)$$

M is non-linear but the tangent linear model R remains, to a large extent, valid for the propagation of the forecast errors (*Lacarra and Talagrand*, 1988; *Vukicevic*, 1991; *Rabier and Courtier*, 1992)

$$\delta x(t+T) = R(t+T,t)\, \delta x(t) + \text{random noise (model errors)}$$

The random noise depicts the fact that the model is not perfect. Here we assume no bias and no time correlation and that we know the covariance matrix Q of the model errors. Relaxation of these assumptions is discussed by *Jazwinski* (1970). At a given time, observations are available

$$y_i = H_i x(t_i) + \text{random noise (observation errors)}$$

We assume known the covariance matrix of observation errors O_i. Then we are in a position to apply an extended Kalman filter.

The forecast step

$$x^f(t_{i+1}) = M(t_{i+1},t_i)\, x^a(t_i) \tag{11}$$

$$B^f(t_{i+1}) = R(t_{i+1},t_i) B^a(t_i) R(t_{i+1},t_i)^t + Q \tag{12}$$

The analysis step

$$x^a(t_i) = x^f(t_i) + K_i(y_i - H_i x^f(t_i)) \tag{13}$$

$$B^a(t_i) = (I - K_i H'_i) B^f(t_i) \tag{14}$$

with

$$K_i = B^f(t_i) H'^t_i (H'_i B^f(t_i) H'^t_i + O_i)^{-1} \tag{15}$$

In the analysis step one recognizes the linear estimation equations (2), (8) and (9). What the Kalman filter brings is a way of making use of the dynamics to transport the information in time through equation (11) and its quality with equation (12).

In the current operational implementation of four-dimensional assimilation relying on OI, (11) is solved using the forecast model. However, solving (12) is an untractable task, as the dimension of the model phase

space is 10^7, (12) would require 10^7 model integrations. (12) is then replaced by a simple evolution law. The correlations are kept constant in time while the variances are assumed to follow a growth according to the typical doubling time of the forecast errors of a couple of days with a saturation toward a climatological value. As a consequence, neither the variances of the forecast errors nor the correlations depend on the meteorological situation. This has been recognized as an OI weakness for years.

4.2 The 4D variational algorithm

4D-Var then consists of the minimization problem

$$\mathcal{P}_{4D}: \text{minimize } J(x(t_o)) = \tfrac{1}{2}(x(t_o)-x_b)^t B^{-1}(x(t_o)-x_b) + \tfrac{1}{2}\sum_{i=0}^{N} (H_i x(t_i)-y_i)^t O_i^{-1}(H_i x(t_i)-y_i) \quad (16)$$

with $x(t_i) = M(t_i,t_o)x(t_o)$

x_b is the background information valid for time t_o which summarises all the information used before time t_o, and B is the error covariance matrix of x_b.

A classical result, assuming a perfect model and linearity of H and M, is that if $x^*(t_o)$ is the result of $\mathcal{P}_{4D}$, then $x^*(t_N) = M(t_N,t_o)\, x^*(t_o)$ can also be obtained applying the Kalman filter to the same statistical estimation problem (*Jazwinski*, 1970; *Ghil et al*, 1981; *Lorenc*, 1986; see *Thépaut and Courtier*, 1991 or *Rabier and Courtier*, 1992 for a detailed presentation using the same notations as here). In meteorological applications, however, H and M are weakly nonlinear: assuming that the tangent-linear operators $\mathfrak{M}$ and H' of respectively M and H satisfy, to acceptable accuracy for meteorological order of magnitude of the estimation error, the relations

$$\begin{aligned} M(t_i,t_o)\,(x(t_o)+\delta x(t_o)) &\simeq M(t_i,t_o)\,x(t_o) + \mathfrak{M}(t_i,t_o)\delta x(t_o) \\ H_i(x(t_i) + \delta x(t_i)) &\simeq H_i x(t_i) + H'\cdot\delta x(t_i) \end{aligned} \quad (17)$$

for perturbation $\delta x(t_o)$, then the 4D-Var problem $\mathcal{P}_{4D}$ is equivalent to the so-called extended Kalman filter under the above-mentioned hypothesis, namely the quasi-linearity, the perfect model and the fixed lag (see previous references). This consists of two steps (^{f} and ^{a} denote respectively forecast and analysis)

4D-Var implicitly uses flow-dependent structure functions as can be seen from equation (12) (*Thépaut et al* 1993a), so that 4D-Var is a scientific improvement on the current operational implementation. Moreover, 4D-Var is also an algorithmic improvement on the Kalman filter (11-15) where the equation (12) has to be solved explicitly, instead of implicitly in 4D-Var.

There are two main weaknesses in the 4D-Var implementation. First the model is assumed to be perfect: in Eq. (12) no source terms Q are present (*Talagrand*, 1988; *Cohn and Parrish*, 1991; *Daley*, 1991;

Wergen, 1992), nevertheless *Derber* (1989) and *Zupanski* (1993) demonstrated how to address a model bias in 4D-Var. Secondly, we do not have access to the analysis error covariance $B^a(t_N)$. Here we suggest that if (12) is approximate anyway, it is not scientifically worthwhile solving it exactly. This idea has been followed by most of the NWP centres which implemented optimal interpolation; the dynamics $\mathfrak{R}$ was replaced by the identity or a simple law for the temporal evolution of the variances. In the Kalman filter context, this is discussed e.g. by *Dee* (1991) and *Cohn* (1992). In other words, it may be scientifically acceptable to replace $\mathfrak{R}$ by an approximate tangent-linear model in (2) provided this approximation is smaller than the approximation of neglecting the model error source term Q.

4.3 The incremental formulation of variational assimilation

Let us assume from now on that $\mathfrak{R}$ is any linear operator, for which we will later stipulate the link with the model M. We define the 4D-Var problem:

$$\wp'_{4D}\text{: minimise } J(\delta x(t_o)) = \tfrac{1}{2}\delta x(t_o)^t\, B^{-1}\, \delta x(t_o) + \tfrac{1}{2}\sum_{i=0}^{N} (H_i x(t_i)-y_i)^t\, O_i^{-1}(H_i x(t_i)-y_i) \qquad (18)$$

with $x(t_i) = M(t_i,t_o)\, x_b + \mathfrak{R}(t_i,t_o)\, \delta x(t_o)$

(we then have $\delta x(t_o) = x(t_o)-x_b$)

remark 1 If $\mathfrak{R}$ is the tangent-linear model, $\wp'_{4D}$ and $\wp_{4D}$ are equivalent to within the accuracy of the tangent-linear approximation.

remark 2 If $\mathfrak{R}$ is any linear operator which we assume would describe exactly the forecast error evolution, $\wp'_{4D}$ leads to the same result as the Kalman filter described by equations (11-15). However, in a nonlinear problem there is no linear operator which describes the error evolution exactly; introducing $\mathfrak{R}$ will then remain an approximation.

remark 3 $\wp'_{4D}$ is better than $\wp_{4D}$ as far as an operational implementation is concerned since we keep the original model M for propagating in time the state of the atmosphere, but use an approximate propagation in time of the errors, thus introducing some flexibility on the cost of 4D-Var.

remark 4 A variant of $\wp'_{4D}$ is the quadratic problem

$$\wp''_{4D}\text{: minimize } J(\delta x(t_o)) = \tfrac{1}{2}\delta x(t_o)^t\, B^{-1}\delta x(t_o) + \tfrac{1}{2}\sum_{i=o}^{N} (y_{b,i}+H'_i\delta x(t_i)-y_i)^t\, O_i^{-1}\, (y_{b,i}+H'_i\delta x(t_i)-y_i) \qquad (19)$$

with $\delta x(t_i) = \mathfrak{R}(t_i,t_o)\delta x(t_o)$

and $y_{b,i} = H[M(t_i,t_o)x_b]$

The cost of $\mathcal{P}'_{4D}$ and $\mathcal{P}''_{4D}$ are similar but the storage requirement for the background trajectory is different: in $\mathcal{P}'_{4D}$ it is the background vertical column at the observation point and in $\mathcal{P}''_{4D}$ it is the observation equivalent of the background which have to be stored. In $\mathcal{P}'_{4D}$, $\mathfrak{R}$ is an approximate linearisation of M, similarly in $\mathcal{P}''_{4D}$, H'_i is an approximate linearisation of H_i. Approximating the full problem by a quadratic one has theoretical advantages since the solution involves solving only linear equations in principle, as is guaranteed unique.

The structure functions used in the current operational at ECMWF T213 optimal interpolation have a cut-off at wave number 63 (*Lönnberg*, 1988). If we were to use a T106 truncation for $\mathfrak{R}$, this would already be an enhancement in terms of resolution. An adiabatic version for $\mathfrak{R}$ with some basic simplified diabatic processes like horizontal and vertical diffusion and surface friction would produce the same benefits in terms of implicit flow dependent structure functions as obtained by *Thépaut et al* (1993).

The CPU cost of an adiabatic semi-Lagrangian T106 L31 model is typically 1/16 of the CPU cost of the T213 L31 version. The gain of the incremental approach is then of one order of magnitude.

remark 5 Another variant of $\mathcal{P}'_{4D}$ can be obtained by replacing $H'_i\delta x(t_i)$ in Eq. 19 by a finite difference:

$$\mathcal{P}'''_{4D} \text{ minimise } J(\delta x(t_o)) = \tfrac{1}{2}\, \delta x(t_o)^t\, B^{-1}\, \delta x(t_o)$$

$$+ \tfrac{1}{2} \sum_{i=o}^{N} (y_{b,i}-\tilde{y}_{b,i}+\tilde{y}_i-y_i)^t\, O_i^{-1} (y_{b,i}-\tilde{y}_{b,i}+\tilde{y}_i-y_i)^t \qquad (20)$$

with $\tilde{y}_{b,i} = H_i(\tilde{M}(\tilde{x}_b(t_o))$ being the model equivalent of the observation obtained from a simplified (low resolution, adiabatic) background trajectory and $\tilde{y}_i$ being the model equivalent of the observation obtained from the simplified trajectory issued from $\tilde{x}_b(t_o) + \delta x(t_o)$. $\tilde{x}_b(t_o)$ is the background but at lower resolution. Using the Taylor formula, one has to first order $\tilde{y}_i - \tilde{y}_{bi} = H'_i Jx(t_i)$. The practical advantage of this formulation over $\mathcal{P}''_{4D}$ or $\mathcal{P}'_{4D}$ is that there is less technical development required once the full 4D-Var problem has already been implemented. It is with this formulation that the numerical experimentations are performed at ECMWF. These three implementations $\mathcal{P}'_{4D}$, $\mathcal{P}''_{4D}$ and $\mathcal{P}'''_{4D}$ are equivalent in the quasi-linear context. They are expected to behave differently in the presence of strong nonlinearities, however we have not seen any arguments as to why one should be superior to the others.

4.4 Further developments

There are two ways of improving the model $\mathfrak{R}$. Firstly one could increase the horizontal resolution: the main drawback here is the cost involved since the CPU follows a power law close to 3. In addition the trajectory storage and thus the IO also follow a cubic law (quadratic at a given time step but the number of time steps increases linearly).

Secondly, it is necessary to take into account the physics. The experiments performed so far (*Thépaut et al*, 1993a and *Rabier and Courtier*, 1992) have used only horizontal and/or vertical diffusion with a simple surface friction. *Rabier et al*, 1993 showed that large-scale condensation is essential in order to get reasonable humidity fields in the upper troposphere. More generally, it is expected that the important feedback loops present in the model $\boldsymbol{M}$ will have to be described to a reasonable accuracy with $\mathfrak{R}$; this is expected to be of particular importance in the tropics. *Zou et al* (1993) and *D Zupanski* (1993) have performed feasibility studies using the adjoint of physical parametrizations. The automatic methods developed at INRIA will assist us in formulating a series of tangent-linear models including progressively more effects of the physics (*Rostaing et al*, 1993).

In terms of cost this will eventually double the CPU cost of 4D-Var (as the cost of the physical parametrizations is about 50% of the cost of the model) but it will immediately double the storage required for the trajectory (and the related IO). Currently, only t values are stored since the dynamics are nonlinear only with respect to these t values and not the $t - \Delta t$. Since, the physics are nonlinear with respect to $t - \Delta t$ values, they too will have to be stored. It should be pointed out, however, that a 2 time-level semi-Lagrangian scheme would not require this extra storage.

The physics is far more nonlinear than the dynamics. As a consequence, the tangent-linear approximation is likely to be less valid for the full model than for the adiabatic version. This means that $\mathcal{P}'_{4D}$ or $\mathcal{P}''_{4D}$ are not necessarily a very good approximation of $\mathcal{P}_{4D}$. A simple way for accounting some of the nonlinearities in the final analysis is to define a sequence $\mathcal{P}''_{4D}(n)$ of assimilation:

$\mathcal{P}''_{4D}(n)$: minimise $J(\delta x^n(t_o)) =$

$$\tfrac{1}{2}\,(\delta x^n(t_o) + x^{n-1}-x_b)^t\, B^{-1}\,(\delta x^n(t_o) + x^{n-1}-x_b) + \tfrac{1}{2}\sum_{i=o}^{N}(y_i^{n-1}+H_i'\,\delta x^n(t_i)-y_i)^t\, O_i^{-1}(y_i^{n-1}+H_i'\,\delta x^n(t_i)-y_i) \qquad (21)$$

with $\quad \delta x(t_i) = \mathfrak{R}(t_i,t_o)\delta x(t_o)$ (22)

and $\quad y_i^{n-1} = H_i[M(t_i,t_o)\,(x^{n-1}+\delta^{*n-1}x(t_o))]$ (23)

$\delta^{*n-1}x(t_o)$ is the result of the (approximate) minimization of $\mathcal{P}''_{4D}(n-1)$ and $\delta^{*o}x(t_o) = 0$

and $x^{n-1} = x^{n-2} + \delta^{*n-1}x(t_o)$ $\qquad$ $x^o = x_b$, $x^{-1} = x_b$

This algorithm can be seen as a pair of nested loops. The outer loop uses the complete model in Eq. (23) to re-define the model trajectory at each iteration of the outer loop. The inner loop uses the tangent-linear and adjoint of a simpler (e.g. adiabatic) model (Eq. 22) to minimize the cost function (Eq. 21) for the increments calculated with respect to the re-defined trajectory.

This approach allows a progressive inclusion of physical processes without dealing with large-scale non-differentiable minimization problems, of which little is known in practice. The drawback is that we have no guarantee that the sequence $\delta^{*n}x(t_o)$ will converge. Experimental work is necessary to address this issue but we have to be pragmatic. Highly non regular problems will remain intractable for a long time but we have here a reasonable approach that is probably robust.

Remark $\mathfrak{R}$ does not have to be kept constant in this iterative process and one can imagine a sequence $\mathfrak{R}^n$ where the resolution and the number of physical processes dealt with increase with n.

4.5 Cycling 4D-Var

As we said earlier, 4D-Var is limited to a fixed lag: a time interval of e.g. 24 hours. It is then necessary to account for past information: cycling 4D-Var. Fig. 1 presents a synopsis of the 4D-Var incremental algorithm. First the background trajectory is compared with the observations producing the integration vector d_i. Then the incremental 4D-Var minimization problem is solved. The minimization is performed for a simplified problem. Introducing ψ, the simplification operator, we introduce the control variable of the simplified problem

$$z_o = \psi \delta x(t_o)$$

Integrating in time the simplified dynamics produces the simplified trajectory z_i. G_i is the simplified observation operator (it is an approximation of $H':\psi^{-1}$). 4D-Var incremental becomes

$$J(z_o) = \tfrac{1}{2}\, z_o^t\, B^{-1}\, z_o + \tfrac{1}{2}\sum_i (d_i+G_i\, z_i)^t\, O_i^{-1}(d_i+G_i\, z_i)$$

The covariance matrix of background error B is specified from statistics. The analysis increments (result of the minimization) are added to the background at time t_o and transported at time t_n using the high resolution model. It is then possible to start the next assimilation cycle.

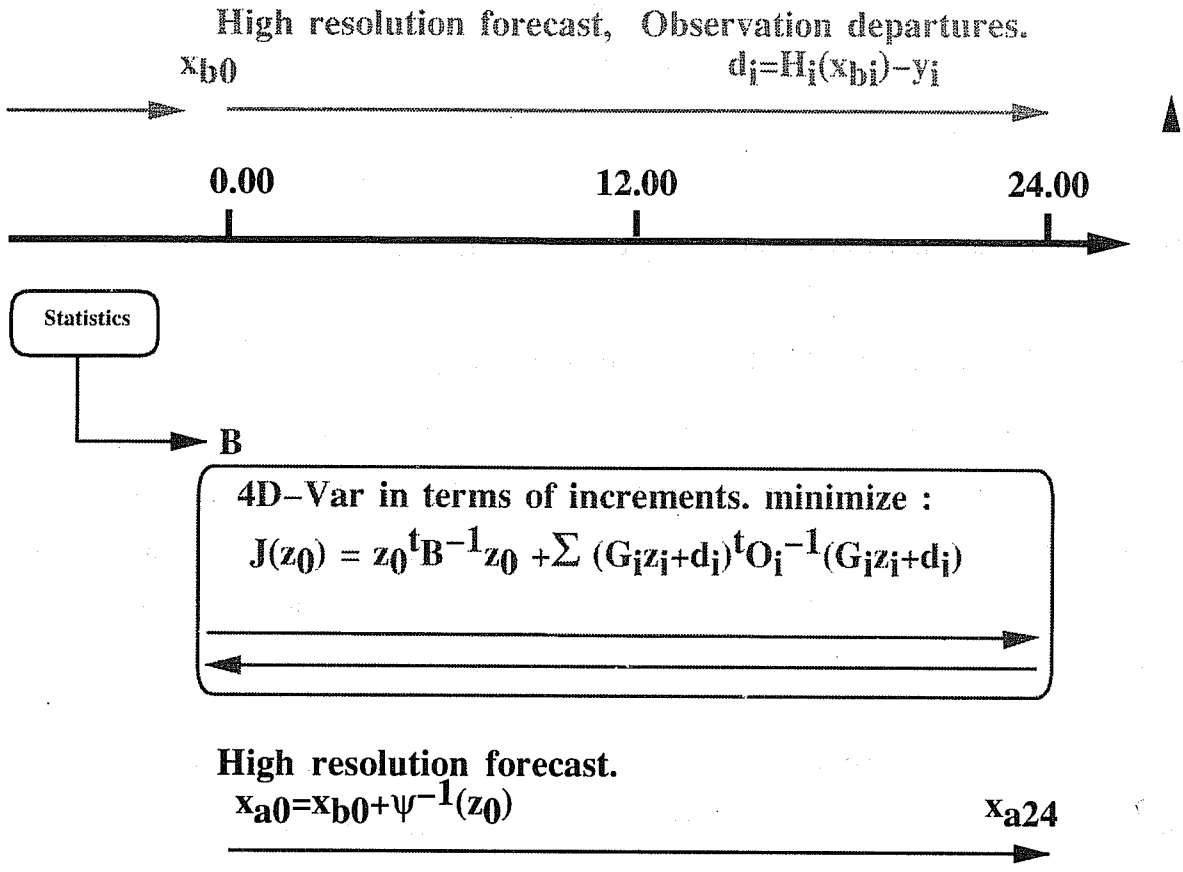

Fig. 1 Synopsis of 4D-Var incremental.

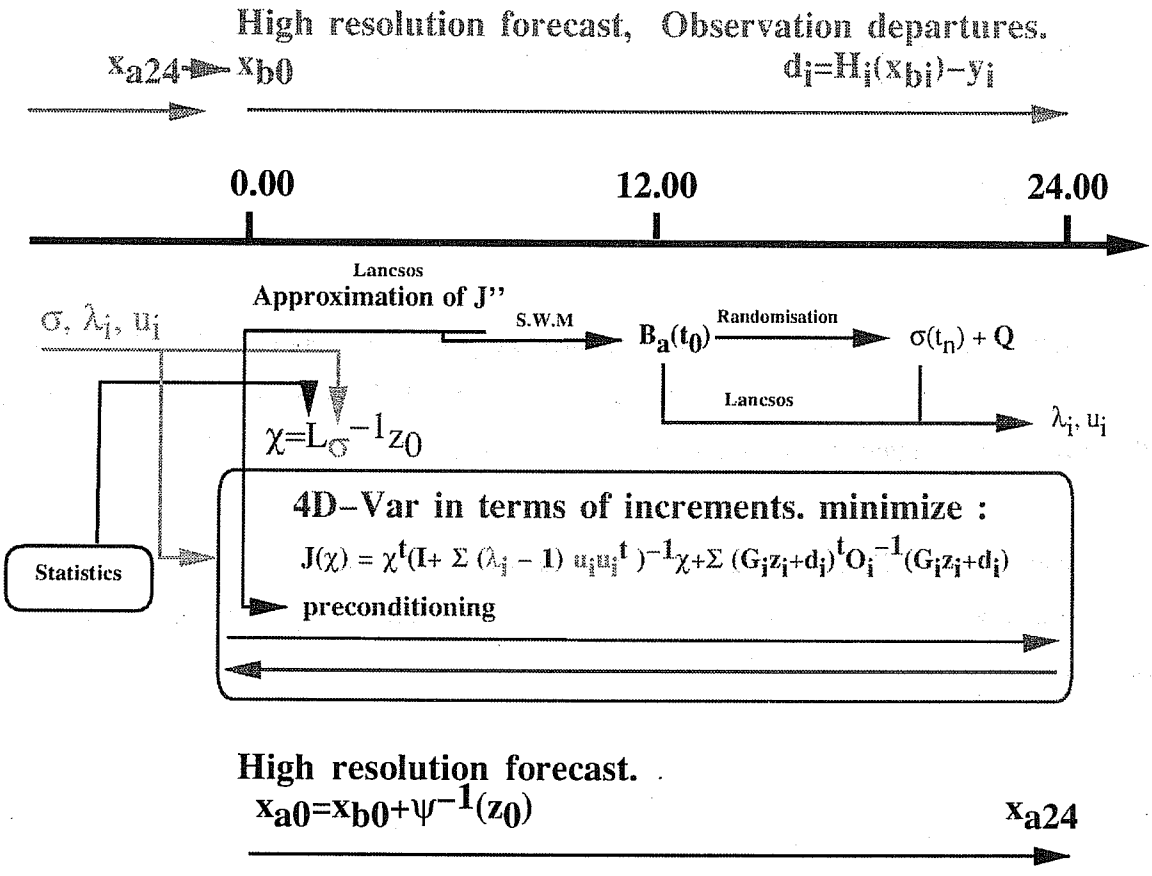

Fig. 2 Synopsis of the proposed algorithm for cycling 4D-Var.

This is very similar to the current operational way of cycling OI. However, one would like to have the standard deviations σ_b of section 3.3 unfixed, but reflecting the previous 4D-Var assimilation quality and, in particular, the data coverage. In the current experimental implementation of 4D-Var at ECMWF, an OI is run in order to prepare and quality-control the observations; in 4D-Var we then use the σ_b used in OI. However, this is not very satisfactory and we now present an algorithm which may overcome this weakness.

As described in *Courtier et al* (1993), $\boldsymbol{B}$ is rewritten as $\boldsymbol{B} = L_\sigma L_\sigma^t$. A change of variable $\chi = L_\sigma^{-1} z_o$ is introduced. The subscript σ denotes the dependence of $\boldsymbol{B}$ on the standard deviation of background error specified in grid point space. The control variable used for the minimization is χ. This is a preconditioning by the Hessian of the background term. Assuming the observations operators linear, J has the same Hessian as a cost function in which the observation departures are set to 0. The advantage of the latter problem being that we now know its minimum: $\chi = 0$.

As described in *Courtier et al* (1993), where it has been applied to simulated observations or in *Rabier and Courtier* (1992), it is then possible to use randomization for the observation term J_o in order to compute an approximation of J''_o. Here we propose to use a Lanczos algorithm. As we know the minimum (0) of the parallel problem, each computation of a gradient in a given direction in phase space provides the Hessian applied to this direction allowing us to implement the Lanczos algorithm without introducing roundoff errors due to finite differences. The Hessian of J_o is thus approximated by a few eigenvectors

$$J''_o = \sum_j v_j v_j^t = vv^t$$

and $$J'' = I + vv^t = \{I - v(I+v^tv)^{-1}v^t\}^{-1} \qquad (24)$$

(24) makes use of the Sherman-Woodbury-Morrison formulae and provides an approximation of the Hessian and of its inverse. This has two important consequences. First, it is possible to precondition the minimization problem accounting for the dominant J_o contribution.

Secondly, the inverse of the Hessian is the covariance matrix of analysis error. The approximation introduced consists of considering the reduction of the estimation error brought by the observations only in the leading directions.

Remark 1 In practice, because of the way we have implemented the multivariate J_b, its Hessian is not exactly the identity. It remains, however, close. The main departure for identity would easily be estimated with a Lanczos algorithm.

Remark 2 Assuming that we would need 100 iterations of the Lanczos algorithm, it is necessary to implement the algorithm for a further simplified problem, e.g. T21 resolution, in order to keep the overcost of order one iteration of minimization of 4D-Var incremental.

The standard deviation of analysis error is obtained in grid point space using randomization (by construction we know the eigenvectors of the approximate Hessian). It is then possible to implement the current way of cycling OI or 3D-Var. However, it is clearly possible to do far better than this.

First the standard deviations at time t_N may easily be computed by randomization. We can then add the model noise Q either with a white or red spectrum. We then have the σ_{t_N} to be used in the definition of the L_σ operators of the following cycle.

Secondly we have

$$B_a(t_N) = L_{\sigma_{tN}}^{-1} R(t_N,t_o) L_\sigma B_a(t_o) L_\sigma^t R(t_N,t_o)^t L_{\sigma_{tN}}^{-1^t}$$

Using a Lanczos algorithm it is possible to approximate $B_a(t_N)$ with a matrix of the form $I + \sum_i (\lambda_i - 1) u_i u_i^t$.

It is thus possible to account for this matrix in the background term thus introducing flow dependency in its specification. The synopsis of the algorithm is depicted in Fig. 2.

CONCLUSION

Using the linear estimation theory, we have introduced the algorithm used operationally in numerical weather prediction. The time dimension has been introduced by comparison with the Kalman filter. The 4D-Var algorithm allows a 4D analysis but on a finite time interval. Section 4.5 presents an algorithm not yet tested to cycle 4D-Var. It is an approximation of the Kalman filter which brings flow dependency of the standard deviation of background error and also of the structure functions, but at lower resolution.

ACKNOWLEDGEMENTS

I would like to thank the ECMWF Data Division, Olivier Talagrand and Jean Pailleux for fruitful discussions. Section 4.5 relies on theoretical developments made in the last three years with Pierre Gauthier, Florence Rabier and Jean-Noël Thépaut. A key idea has been provided by Jorge Nocedal. Carole Edis typed the manuscript with a lot of patience.

REFERENCES

Boer, G J, 1983: Homogeneous and isotropic turbulence on the sphere. J Atmos Sci, 40, 154-163.

Cohn, S E and D F Parrish, 1991: The behaviour of forecast error covariances for a Kalman filter in two dimensions. Mon Wea Rev, 129, 1757-1785.

Cohn, S E, 1992: Short-term dynamics of forecast error covariances. Proceedings of the ECMWF workshop on "variational assimilation with special emphasis on three-dimensional aspects", 9-12 November 1992, Reading, UK, pp 157-170.

Courtier, P, J-N Thépaut and A Hollingsworth, 1993: A strategy for operational implementation of 4D-Var using an incremental approach. Q J R Meteorol Soc (submitted).

Daley, R, 1991: Atmospheric data analysis. Cambridge atmospheric and space science series. Cambridge University Press.

Dee, D P, 1991: Simplification of the Kalman filter for meteorological data assimilation. Q J Meteorol Soc, 117, 365-384.

Derber, J C, 1989: A variational continuous assimilation technique. Mon Wea Rev, 117, 2437-2446. 710.

Durand, Y, 1985: The use of satellite data in the French high resolution analysis. Proceedings of the 1985 ECMWF workshop on high resolution analysis, 24-26 June 1985. pp 89-128.

Gandin, L S, 1963: Objective analysis of meteorological fields translated from Russian by the Israeli Programme for Scientific Translations, 1965.

Ghil, M, S Cohn, J Tavantzis, K Bube and E Isaacson, 1981: Application of estimation theory to numerical weather prediction. Dynamic meteorology: data assimilation methods, L Bengtsson, M Ghil and E Källen, eds, Springer Verlag, 330 pp.

Ghil, M and P Manalotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. Adv Geophys, 33, 141-266.

Hollingsworth, A and P Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. Tellus, 38A, 111-136.

Jaszwinski, A H, 1970: Stochastic processes and filtering theory. Academic Press, New York.

Kalman, R E, 1960: A new approach for linear filtering and prediction problems. Trans ASME, Ser D, J Basic Eng, 82, 35-45.

Lacarra, J F and O Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. Tellus, 40A, 81-95.

Lönnberg, P and A Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. Tellus, 38A, 137-161.

Lönnberg, P, 1988: Developments in the ECMWF analysis system. 1988 ECMWF seminar on data assimilation and the use of satellite data, pp 75-119, Reading, UK.

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

Lorenc, A, 1986: Analysis methods for numerical weather prediction. Q J R Meteorol Soc, 112, 1177-1194.

McGrath, R, 1993: Operational monitoring of satellite data and its availability. Proceedings of the 1993 ECMWF seminar on "Developments in the use of satellite data in Numerical Weather Prediction" 6-10 September 1993.

Parrish, D F and J Derber, 1992: The National Meteorological Centre's spectral statistical interpolation analysis system. Mon Wea Rev, 120, 1747-1763.

Phillips, N A, 1986: The spatial statistics of random geostrophic mode and first-guess error. Tellus, 38A, 314-322.

Rabier, F and P Courtier, 1992: Four-dimensional assimilation in the presence of baroclinic instability. Q J R Meteorol Soc, 118, 649-672.

Rabier, F, P Courtier, J Pailleux, O Talagrand and D Vasiljevic, 1993: A comparison between 4D variational assimilation and simplified sequential estimation relying on 3D variational analysis. Q J R Meteorol Soc, 119, 845-880.

Rabier, F and A McNally, 1993: Evaluation of forecast error covariance matrix. Tech Memo 195, ECMWF.

Rostaing, N, S Dalmas and A Galligo, 1993: Automatic differentiation in odyssée. To appear in Tellus.

Shaw, D B, P Lönnberg, A Hollingsworth and P Undén, 1987: The 1984/1985 revisions of the ECMWF mass and wind analysis. Q J R Meteorol Soc, 113, 553-566.

Simmons, A J, 1991: Development of a high resolution, semi-Lagrangian version of the ECMWF forecast model. Proceedings of the 1991 ECMWF seminar on "Numerical methods in atmospheric models" 9-13 September 1991.

Talagrand, O, 1988: Four-dimensional variational assimilation. Proceedings of the 1988 ECMWF seminar, Reading, UK.

Tarantola, A, 1987: Inverse problem theory. Methods for data fitting and model parameter estimation. Elsevier.

Thépaut, J-N and P Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of a multi-level primitive equation model. Q J R Meteorol Soc, 117, 1225-1254.

Thépaut, J-N, R Hoffman and P Courtier, 1993: Interactions of dynamics and observations in a four-dimensional variational assimilation. Mon Wea Rev, 121, 3393-3414.

Vukicevic, T, 1991: Non linear and linear evolution of initial forecast errors. Mon Wea Rev, 119, 1602-1611.

Wergen, W, 1992: The effect of model errors in variational assimilation. Tellus, 44A, 297-313.

Zou, X, I M Navon and J Sela, 1993: Variational data assimilation with moist threshold process using NMC spectral model. Tellus in press.

Zupanski, D, 1993: The effect of discontinuities in the Betts-Miller cumulus convection scheme on four-dimensional data assimilation. Tellus in press.

Zupanski, M, 1993: Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. Mon Wea Rev, 121, 2396-2408.