TECHNICAL MEMORANDUM

210

Quasi continuous variational data assimilation

H. Järvinen, J N. Thépaut and P. Courtier

Research Department

January 1995

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

ABSTRACT

All the computations associated to the operational meteorological data assimilation are usually carried out after the so called cut-off-time, i.e. a few hours after the assimilation period. In this article we propose an alternative implementation of the four-dimensional variational assimilation where some of the computations occur during the 24-hour assimilation period. The feasibility of the approach is first validated with an assimilation system built around a low resolution barotropic grid point model. The results are then confirmed with a multi-level primitive equation model with real observations. The main result of these experiments is that the peak computer power requirement of the four-dimensional variational data assimilation may be significantly reduced by the suggested approach.

1. INTRODUCTION

Numerical weather prediction (NWP) is facing a major change as the Optimum Interpolation (OI) analysis method (*Gandin*, 1963) is gradually giving way to the variational assimilation: the techniques to extract observational information and to provide initial conditions for deterministic prediction models are changing (*Parrish and Derber*, 1992). The implementations of OI imply choices that are necessary in the OI-environment but which may be relaxed in a different environment. One such constraint related to the OI is discussed in this article and an alternative solution to implement the variational assimilation will be suggested.

At present, the most widely used data assimilation method is OI which provides an estimate of the analysis value at model grid points by constructing a statistically optimal linear combination of observations. That is achieved (e.g. *Lorenc*, 1986) through

$$\boldsymbol{x}_a = \boldsymbol{x}_b + \boldsymbol{B}\boldsymbol{H}^t (\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^t + \boldsymbol{O})^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}_b) \qquad (1)$$

where $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ are the vectors containing the analysis and the background values, respectively. The observations, making up the vector $\boldsymbol{y}$, are related to $\boldsymbol{x}_b$ through the linear observation operator $\boldsymbol{H}$ which defines the components of $\boldsymbol{x}_b$ that has been observed. The observation errors enter through the covariance matrix $\boldsymbol{O}$ and the uncertainty associated with $\boldsymbol{x}_b$ is described by the short-range forecast error covariance matrix $\boldsymbol{B}$, as $\boldsymbol{x}_b$ usually is a 6-hour forecast from the previous analysis. Matrix inversion and transpose are denoted by $^{-1}$ and ^{t}, respectively.

The matrix inversion involved in (1) is an expensive operation and the amount of observations that can be handled simultaneously in OI has to be limited. At the European Centre for Medium-range Weather Forecasts (ECMWF) that dimension limit is currently 621. Because of this restriction, the analysis calculation is made for boxes where a small areal subset of the observations is used in one time. The global fields are formed by blending adjacent boxes together and thus most of the available observations can be used in the OI- analysis (*Lorenc*, 1981; *Shaw et al*, 1987).

Despite of the areal division of the observations to overcome the matrix size restriction, a limitation concerning the operational implementation of OI lies just here. Although in theory OI can be formulated as a sequential algorithm (*Jazwinski*, 1970; *Courtier*, 1987), in practice it remains as a static assimilation method. Therefore, once the matrix inversion in (1) is completed, the analysis cannot be influenced any more. Consequently, in OI it is of paramount importance to choose the best possible set of observations to enter the analysis before the matrix inversion. The operational implementation of OI is thus constrained, as the best way to choose the areal subset of observations is to wait until the end of the period from which the observations are collected. The assimilation of the observations can thus begin only after a cut-off-time - from 3 to 12 hours in different applications - after the end of the current assimilation period. This allows delayed observations (e.g. by slow telecommunication) to be used in the assimilation process.

Most operational NWP centres have indeed adopted a sequential algorithm for data assimilation and forecasting. This implies a concentrated computing demand just after the cut-off-time, when the assimilation requires peak computer power (Fig 1, uppermost panel).

In the following, the focus is on the four-dimensional variational data assimilation (4D-Var) which consists essentially of the minimization of a cost function J (e.g. *Courtier et al*, 1993)

$$J(x(t_0)) = \frac{1}{2}(x(t_0)-x_b)^t B^{-1}(x(t_0)-x_b) + \frac{1}{2}\sum_{i=0}^{T}(H_i x(t_i)-y_i)^t O_i^{-1}(H_i x(t_i)-y_i) \qquad (2)$$

Here $x(t_0)$ denotes the model initial state at the beginning of the assimilation period T, i.e. the control variable of the problem, whereas the following states $x(t_i)$ at time i are obtained through the time integration of the forecasting model. The observation operator H_i can be a simple interpolation function if an observation is made of a model variable but it can be a complicated routine, eg. a radiative transfer model in the case of radiance measurements. The subscript $_i$ in the observation error covariance matrix O_i (omitted hereafter) states that the matrix O is dependent on the observations available at the time i.

The formulation (2) is chosen so that when the time dimension in (2) is not considered and H is linear, x_a of (1) would minimize the cost function; this shows the equivalence of OI and three-dimensional variational data assimilation. Equation (1), on the other hand, corresponds to the analysis step of the Kalman filter and minimizing (2) leads to the same solution that would be obtained with the Kalman-filter in the case of linear dynamics and assuming a perfect model, resulting in the best estimate of the atmosphere at the end of the assimilation period (e.g. *Rabier et al*, 1993).

Equation (2) can be written symbolically as a sum

$$J = J_b + J_o = J_b + \sum_{i=0}^{T} J_{oi} \qquad (3)$$

where J_b and J_o are the background and the observation terms, respectively. J_b measures the misfit between the model initial state at $t = t_0$ and all available information from before the assimilation period, summarized by the background field x_b. J_o measures the distance of the model state to the observations at appropriate times during the assimilation period. The term J_o consists of several individual terms J_{oi} corresponding to short time slots of the assimilation period T and can be written as a sum as in (3). The cost function can also include some dynamical or physical constraints or a measure of distance to other possible source of information in form of an additional weak constraint term J_c in sum (3). In the present study, term J_c is a penalty for the presence of gravity waves which are measured by the distance of model initial state to the slow manifold (*Courtier and Talagrand*, 1990). For more details on the formulation of the cost function see e.g. *Heckley et al* (1992) and *Vasiljevic et al* (1992).

The formulation of the minimization problem implies an assimilation period T over which the model is integrated and the comparison with the available observational information is made. The methods currently envisaged to implement 4D-Var in an operational environment (e.g. *Pailleux*, 1989 and 1993; *Courtier et al*, 1993 and 1994; *Derber et al*, 1992; *Rabier et al*, 1993) are based on a configuration similar to those used in OI implementations: the assimilation of the observations starts after the end of assimilation period, when the cut-off-time is reached (Fig. 1, middle panel). In 4D-Var the assimilation period T can be as long as 24 hours whereas in OI a typical assimilation period is six hours. The amount of observations may therefore be large in 4D-Var compared with OI.

The calculation and minimization of the cost function is a computer intensive task. The main limitation for the application of full variational approach is the requirement for peak computer power. Methods for better preconditioning of the problem (*Courtier et al*, 1994) and for an incremental approach (*Derber et al*, 1992; *Courtier et al*, 1993 and 1994) have been proposed to reduce the computational cost. The former allows the solution of the assimilation to be achieved with a smaller computational effort. The latter is an approximation of the full problem and allows one to regulate the time used for the assimilation and the accuracy of the solution, depending on how much the linear model is simplified for the minimization. It also provides a way of including certain physical processes into the 4D-Var without developing the adjoint of the full forecast model.

2. PARTITIONING OF THE COMPUTATIONS OVER THE ASSIMILATION PERIOD

We propose an alternative approach to the implementation of variational assimilation by relaxing the constraint related to the completeness of data, keeping in mind that the real limitation for the application of 4D-Var is the peak computer power. We envisage a quasi-continuous process of data assimilation which exploits the inherent freedom of 4D-Var to begin the assimilation with an incomplete set of observations. The assimilation is started before all the observations are available and the assimilation task is distributed

in time. Although the set of observations is incomplete, the assimilation is still global. The aim is to reduce the minimization task that otherwise should be completed in one calculation.

2.1 Basic principles of the approach

Let us divide the assimilation period T into N sub-windows, the duration of each being T/N. The length of T may be 24 hours and N is a suitable integer, e.g. 4.

The first time-partitioned assimilation (A_1) is done with the observations that have arrived during and by the end of the first sub-window. There is thus available at $t = T/N$ a fraction $1/N$ of the total number of observations of the assimilation T period if a constant observation flow is assumed. These observations will be preprocessed after a suitable spell of time (or sub-cut-off-time) shortly after $t = T/N$. The background information is available in the form of $\boldsymbol{x}_b$ valid at the beginning of the assimilation period $t = t_0$. The vector $\boldsymbol{x}_b$ may also be used as the initial point of minimization as it has an optimal fit to observations just before the current assimilation period. The background and observational error covariance matrices, $\boldsymbol{B}$ and $\boldsymbol{O}$, that are needed in the calculation of the cost function are the usual ones used in 4D-Var (*Courtier et al*, 1993). This first assimilation is identical to the normal 4D-Var except that the assimilation period is T/N instead of T. After completing the first assimilation task, the result of the minimization is stored in the form of a model field valid at $t = t_0$.

At the end of the second sub-window, the amount of observations has increased by roughly a fraction $1/N$. There may also be observations that logically belong to the first sub-window but which arrived too late to be used in assimilation A_1. These can be included into the second assimilation process (A_2) which extends over two consecutive windows from $t = t_0$ to $t = 2T/N$ using all the observations that have arrived by $t = 2T/N$. The background information, $\boldsymbol{x}_b$ and $\boldsymbol{B}$, is unchanged whereas the initial point of this minimization is the result of the assimilation A_1 which fitted the observations that arrived between $t = t_0$ and $t = T/N$ and should be a better approximation of the atmosphere than $\boldsymbol{x}_b$.

The quasi-continuous variational data assimilation proceeds in this way by assimilating the available observations already during the assimilation period and improving repeatedly the model initial state. The assimilation period increases each time by a period of T/N. Comparing the normal and quasi-continuous 4D-Var at the end of T when the last assimilation is beginning, the difference is that the model initial state, or the estimate for the analysis at $t = t_0$, has undergone $N-1$ iterative improvements. Fig 1 (lowermost panel) shows schematically how the computations proceed in the quasi-continuous approach compared with a normal 4D-Var implementation (Fig 1, middle panel). In the quasi-continuous approach there are no strict scheduling requirements for the assimilation tasks other than the logical order and the requirement that the last assimilation must finish before the time-critical forecast job can be started.

In the normal implementation of 4D-Var, where the minimization is performed in one calculation over $\boldsymbol{T}$, the cost function $\boldsymbol{J}$ can be written according to (3) in the case $N = 4$ as follows

$$J = J_b + J_o = J_b + J_{o1} + J_{o2} + J_{o3} + J_{o4} \tag{4}$$

The quasi-continuous approach $\boldsymbol{A}$ consists of a set of consecutive assimilation tasks with a different cost function to be minimized each time. These are

$$\begin{aligned} A_1&: \quad J = J_b + J_{o1} \\ A_2&: \quad J = J_b + J_{o1} + J_{o2} \\ A_3&: \quad J = J_b + J_{o1} + J_{o2} + J_{o3} \\ A_4&: \quad J = J_b + J_{o1} + J_{o2} + J_{o3} + J_{o4} \end{aligned} \tag{5}$$

The initial point of minimization A_i is always the result of the previous minimization task A_{i-1}. This algorithm actually bears a similarity with the Kalman filter where the background is combined repeatedly with the observations to produce an updated background (e.g. *Lorenc*, 1986). In (5) the background is also combined with the latest observations but to form a new initial point of minimization through the iterative search. The background itself remains unchanged throughout the assimilation.

If we assume a stable observational network, the formulation (5) suggests a way to estimate all the observation terms J_{oi} by those currently available. That can be done by taking into account the number of observations available compared to the total amount and weighting accordingly. An alternative algorithm $\boldsymbol{B}$ can then be written in the case $N = 4$ as

$$\begin{aligned} B_1&: \quad J = J_b + 4J_{o1} \\ B_2&: \quad J = J_b + 2(J_{o1} + J_{o2}) \\ B_3&: \quad J = J_b + \frac{4}{3}(J_{o1} + J_{o2} + J_{o3}) \\ B_4&: \quad J = J_b + J_{o1} + J_{o2} + J_{o3} + J_{o4} \end{aligned} \tag{5}$$

This algorithm is sub-optimal in the sense that it is not equivalent to the optimal Kalman filter, i.e. in (6) the *ad hoc* weights given to the observations do not appear in the Kalman filter. Numerical results show the algorithm (5) to be superior to (6) and therefore (6) will not be developed further. The quasi-continuous approach will refer in the following to the algorithm (5).

2.2 Preconditioning of the minimization

The quasi-continuous approach allows the possibility of gathering preconditioning information of the minimization during the assimilation process. In the quasi-Newtonian iterative search of the minimum of

cost function J (*Gilbert and Lemaréchal*, 1989), the nth estimate of the control variable x is updated through the formula

$$x_{n+1} = x_n - \overline{\left(\left[\nabla^2_{x_n} J(x_n)\right]^{-1}\right)} \cdot \nabla_{x_n} J(x_n) \tag{7}$$

The gradient -vector of the cost function with respect to the control variable, $\nabla_x J$, is obtained through integration of the adjoint model. The dimension of the square matrix containing the second partial derivatives, the Hessian $\nabla_x^2 J(x)$, equals the dimension of the control variable and would be the optimal preconditioning of the minimization problem. However, the Hessian is unknown and is too large in most meteorological applications to be handled explicitly. Therefore only an approximation, denoted by an overbar, is available in (7). The diagonal of the Hessian provides useful information to determine the optimal descent direction and step length in (7), as reported by *Thépaut and Moll* (1990) who explicitly calculated the Hessian in the case of a low dimensional model.

The gradient -vectors evaluated during the course of the minimization can be used to estimate and to update the diagonal of the Hessian matrix, thereby improving the convergence. In the first few steps no information of this kind is available, if not explicitly provided. One could think of using the Hessian from the previous day, which has been found to improve the convergence (*Courtier*, 1987; *Courtier et al*, 1994). The quasi-continuous approach provides, in principle, a possibility to use the Hessian information of the assimilation A_{i-1} as preconditioning when starting the assimilation A_i. The idea is to estimate the diagonal elements of the Hessian in the first assimilation and pass that information to the next assimilation in the quasi-continuous chain which in turn passes the most recent update to the next one, and so forth. There may be, however, a practical difficulty. The Hessian is given by equation

$$\nabla_x^2 J(x) = B^{-1} + H^t O^{-1} H \tag{8}$$

where B, H and O are as defined in (1). Although the matrix B may be kept constant over the whole assimilation period, O may not, because it is dependent on the observational network. In the quasi-continuous approach the intention is to use a larger set of observations in each successive assimilation and therefore the matrix O will be different and, consequently, also the Hessian. Numerical experiments will show how this aspect of quasi-continuous approach works in practice. One may note that (8) equals the inverse of the analysis error covariance matrix at the beginning of the assimilation period.

2.3 The convergence and the length of assimilation period

The rate of convergence when minimizing J using the adjoint of a primitive equation model varies with the length of the assimilation period and also during the minimization (*Thépaut and Courtier*, 1991). Generally, the shorter the assimilation period, the faster the convergence. *Li et al* (1993) and (1994) draw similar conclusion. The convergence also tends to slow down when approaching the minimum of the cost function as the minimization starts to be saturated. Therefore, the experimentation with the quasi-continuous

approach is planned accordingly. The first assimilation tasks of short assimilation period utilize the associated rapid convergence. By searching only an approximate solution of the minimum of the cost function in the assimilations A_1, A_2 and A_3, the rapid convergence at early stages of minimization is thus also utilized. As the largest scales converge most rapidly (*Thépaut and Courtier*, 1991), terminating the minimization before the saturation implies that the small scale structures have not effectively been generated in the model initial state.

3. FEASIBILITY STUDY WITH A BAROTROPIC MODEL

3.1 The model and construction of variational problem

We use a barotropic vorticity equation model for a feasibility study. The model variable, the 500hPa geopotential height, is given in the National Meteorological Center's (NMC) polar-stereographic grid, at a total of 1404 grid points north of 20°N. The number of degrees of freedom is 1236 as there is no tendency at the boundaries. The grid interval is 381km at latitude 60°N. The model is described in detail in *Rinne and Järvinen* (1993).

The experimental framework uses the model to produce synthetic observations, hereafter simply referred to as *the observations*, distributed along the assimilation period $\boldsymbol{T}$. This is done by choosing randomly one initial condition of the model, hereafter referred as *the truth*, from analyses provided by the NMC. Then a spatially uniform but sparse set of grid point values of forecast fields at full hours are selected as observations $\boldsymbol{y}$, to which random errors are added. The number of observations equals 1236 with 51 or 52 observations at each full hour {1, 2, ..., 24} after the initial time. There is, in other words, exactly one observation in each grid point (excluding the boundary) during the 24-hour assimilation period. The background field $\boldsymbol{x_b}$ is equal to the truth plus random errors added everywhere but on the boundary.

The cost function is constructed from (2) with a few simplifications. $\boldsymbol{B}$ and $\boldsymbol{O}$ are replaced by constant error variances of the background σ_b^2 and of the observations σ_y^2, respectively. The synthetic observations appear at model grid point locations and therefore the observation operator $\boldsymbol{H}$ reduces to a projection. In the filtered model there are no gravity waves and so the term $\boldsymbol{J_c}$ is missing. The cost function $\boldsymbol{J}$ now takes the form

$$J(x(t_0)) = J_b + J_0 = \frac{1}{2}\sum_{i=}^{N} \frac{1}{\sigma_b^2}(x_i(t_0)-x_{bi})^2 dA_i + \frac{1}{2}\sum_{j=1}^{T}\sum_{i=1}^{M} \frac{1}{\sigma_y^2}(x_i(t_j)-y_i(t_j))^2 dA_i \qquad (9)$$

The summations run over the number of grid points (N), over observational times (T) and over the number of observations (M). Here N=M=1236 and T=24. The areal elements dA sum up to unity over the grid and normalize the squared departures at different latitudes to have the same weight in the cost function.

The cost function J is controlled through the model initial state $x(t_0)$, which is the control variable of the problem. For the minimization the partial derivative of J with respect to the control variable $x(t_0)$ is needed. The partial derivative of the first term of (9) for grid point i is simply

$$\left(\frac{\partial J_b}{\partial x(t_0)}\right)_i = \frac{1}{\sigma_b^2}(x_i(t_0) - x_{bi})dA_i \tag{10}$$

This term forces the solution towards the background field and in the absence of observations, $x(t_0) = x_b$ would minimize (9). The second term does not contain $x(t_0)$ but only model variables of later time steps that are related to $x(t_0)$ through the model equation. The term

$$\left(\frac{\partial J_b}{\partial x(t_0)}\right)_i \tag{11}$$

is achieved through integration of the adjoint of the tangent linear model using

$$\left(\frac{\partial J_b}{\partial x(t_0)}\right)_i = \frac{1}{\sigma_y^2}(x_i(t_j) - y_i(t_j))dA_i \tag{12}$$

as the adjoint model variable. The resulting term forces the solution towards the observations. In the absence of the background field x_b the solution minimizing the cost function would have a very close fit to the observations. The development and testing of the adjoint of the barotropic model is described in detail in *Järvinen* (1993) following *Courtier* (1987), *Talagrand* (1991) and *Pailleux et al* (1991).

3.2 Design of the assimilation experiments

The aim of these assimilation experiments is to compare the cost of the quasi-continuous and the normal 4D-Var in terms of computing time. Care was taken to minimize the artificial variation in computing time from one experiment to another. The experiments were performed on a workstation where the main load come from the experiment itself. The computing time fluctuations were small, typically 1-2% between two identical tasks.

In each experiment, the initial point of the minimization is a 12-hour forecast from the truth. A normal 24-hour 4D-Var assimilation is performed starting from this initial point, to find the minimum of the cost function. The minimization is terminated when the decrease of the cost function per iteration becomes smaller than a predefined criterion. The corresponding value of the cost function J, as well as the computing time to reach it, are used as reference values. This first normal 4D-Var assimilation to provide reference values is hereafter called *the control* assimilation.

Next, the quasi-continuous approach is applied. The initial point of the minimization of the first 6-hour assimilation A_1 is the same as in the control case. The convergence of the minimization over this short

assimilation period is very rapid and only an approximate solution of the minimum of the cost function is sought. This improved initial point of minimization, i.e. the control variable after the 6-hour assimilation A_1, is then used as the initial point of minimization in the 12-hour assimilation A_2. An approximate solution of the minimization is again sought. The next assimilation of 18-hours (A_3) further improves the initial point of minimization. The final 24-hour assimilation A_4 uses the latest improved initial point of minimization. In these experiments, the initial value of the cost function of assimilation A_4 was about 1/20 or 1/30 of the corresponding value of the control assimilation. The minimization is terminated as soon as the reference value of the cost function of the control assimilation is reached. This last assimilation A_4 uses most computing time, as only here the actual minimum of the cost function in the quasi-continuous approach is sought.

3.3 Cost comparison of assimilation experiments

Five assimilation experiments were carried out using independent NMC- analyses dated 13 and 25 Dec 1965, 7 and 20 Jan 1966 and 2 Feb 1966, all at 00UTC. Several flow types are therefore covered. The error variance in the background σ_b^2 and in the observations σ_y^2 is 100 m^2. The variational problem (9) is insensitive to this level of error variance as long as it is the same in both terms J_b and J_o. If the error variances were different for the background and observations, the cost of the minimization would also change implying, according to (9), a change in mutual weight of terms J_b and J_o. In the 6-, 12- and 18-hour assimilations (A_1, A_2 and A_3) only 5 iterations are performed with on average a total of 8-10 simulations, i.e. evaluations of the cost function. The minimization is carried out with a quasi-Newtonian conjugate gradient routine of the NAG- library (E04DGF).

The average total cost of performing the quasi-continuous assimilations A_1, A_2, A_3 and A_4 compared to the cost of the 24-hour control assimilation is 64%. The cost of the 24-hour assimilation A_4 alone compared to the cost of the 24-hour control assimilation is 41%. These figures indicate the potential in the quasi-continuous approach.

To verify that the solution of the minimization is the same in the quasi-continuous and control assimilations, we calculated the mean squared difference between the solution and respectively the truth, the background, the initial point of minimization and the observations in both the quasi-continuous and control cases. With an iterative search of the minimum of the cost function, there is an infinite number of possible solutions. The statistics were, however, very close to each other in the quasi-continuous and control assimilations and one could consider the solution to be effectively the same. The difference in computing time needed to achieve these solutions cannot be explained by the difference in the two solutions.

In our first experiment the minimization was terminated in the assimilations A_1, A_2 and A_3 well before the actual minimum of the cost function was reached. We also made an experiment with a single NMC-analysis where the actual minimum is sought in each minimization A_1, A_2, A_3 and A_4. The total computing time of the quasi-continuous assimilation is now longer than that of the 24-hour control assimilation. To study this in more detail, the mean squared difference between the model state and the observations is displayed in Fig 2 over the 24-hour assimilation period in two experiments: the search of minimum is either inaccurate in the 6-, 12- and 18-hour assimilations A_1, A_2 and A_3 (panel a) or accurate (panel b). In the case of an accurate search of the minimum of J, the computing time of the assimilations A_1, A_2 and A_3 was two-fold compared with the case of an inaccurate search. Note that the comparison with observations is made e.g. for the 6-hour assimilation A_1 for the period of 24-hours even though the assimilation itself covers only the first 6-hours. It seems to be that the difference between the model state and the observations is small only over the assimilation period in question, i.e. for A_1 during the first 6 hours, for A_2 during the first 12 hours and so forth. Furthermore, this in only weakly dependent on the accuracy of the minimization. If this had not been the case, one would have expected an improvement of convergence in following assimilation A_{i+1}. Therefore it is beneficial to set a loose criterion to terminate the minimization in the assimilations A_1, A_2 and A_3.

In the barotropic experiments, the computational cost of the quasi-continuous approach is lower than in the control assimilation, provided we perform an approximate minimization in the assimilations A_1, A_2 and A_3. On average, the total cost of the quasi-continuous assimilation compared with the 24-hour control assimilation is 64% and the 24-hour assimilation A_4 costs 41% of the 24-hour control assimilation. No benefit is achieved by saturating the minimization in the assimilations A_1, A_2 and A_3 because of the small scale flow patterns that develop in short assimilation periods. These are not realistic on longer time scales and the system has to get rid of the energy associated with these structures during the subsequent assimilations. It is better to seek an approximate solution before the last assimilation and consider the quasi-continuous approach as a possible preconditioning method of minimization. The following chapter will cast some light on how approximate the solution in assimilations A_1, A_2 and A_3 should be.

4. EXPERIMENTS WITH THE ECMWF 4D-Var ASSIMILATION SYSTEM

The feasibility study with a barotropic model showed that the quasi-continuous 4D-Var is a potentially beneficial approach to the variational data assimilation problem. However, the results of the simplified experiments may not be directly applicable to an operational data assimilation problem. Therefore the evaluation is repeated with an assimilation system of larger dimension and with real observations. In this section we experiment with the ECMWF 4D-Var data assimilation system which can provide a more reliable estimate of the possible benefits of the quasi-continuous approach.

4.1 Description of the assimilation experiments

The atmospheric model used in the experiments is the ECMWF global primitive equation model in adiabatic form truncated at horizontal resolution T21 and discretized in the vertical to 19 levels. There are 28072 degrees of freedom. The only physical processes present in the model are horizontal and vertical diffusion and a simplified surface friction. The adjoint of the model also includes these processes so the model and its adjoint are consistent. The cost function is as defined in (2) with a constraint J_c to control the amount of gravity waves present in the solution (*Thépaut et al*, 1993b).

The set of operational observations is dated 13 Oct 1987 and four sub-sets from 15 to 21UTC, 15 to 03UTC, 15 to 09UTC and 15 to 15UTC (14 Oct 1987), respectively, are created. Table 1 summarizes the number of observations of different types available in the assimilation periods.

		15-21UTC	15-03UTC	15-09UTC	15-15UTC
SYNOP	(u, v, T)	1554	3126	4723	6352
AIREP	(u, v)	1060	2356	3790	5262
SATOB	(u, v)	364	1734	3248	4822
DRIBU	(u, v, T)	297	582	833	1051
TEMP	(u, v, z)	2996	22619	24914	45613
PILOT	(u, v)	2618	4382	6984	8732
Σ		8889	34799	44492	71832

Table 1. The number of observations of different types in the assimilation experiment. The parameters considered are horizontal wind components (u and v), temperature (T) and geopotential height (z). Observation types are denoted by SYNOP for synoptic surface observations, AIREP for aircraft reports, SATOB for geostationary satellite winds, DRIBU for drifting buoys, TEMP for radio-sonde observations and PILOT for wind soundings.

The number of most observation types increase steadily in time, except temp-soundings. This is due to the fact that sounding observations are predominantly carried out at two observation times, 00 and 12 UTC. Furthermore, they are concentrated on continental areas. This spatial and temporal inhomogeneity of the sounding network is significant as they form the main information source among the observations. The largest change in the observational network occurs when going from a 6- to a 12-hour assimilation period as the total number of observations grows by a factor of four. Note that humidity does not appear in Table 1 since the model used is dry. The total number of pieces of information in Table 1 is more than two times the number of degrees of freedom of the T21-model. In all these experiments, the partitioning of the observations is performed *a posteriori*, and therefore the data sets are partial but complete. In the operational data assimilation with the quasi-continuous approach, the data sets may be partial and incomplete due to the irregular arrival of observations.

The assimilation experiment is performed in a similar way as in the barotropic case. First, a 24-hour 4D-Var control assimilation is carried out to provide the reference for the quasi-continuous case. The initial

point of minimization is the background x_b which is a short range (6-hour) forecast. An essential point of the control assimilation is to determine the degree of convergence which corresponds to an acceptable solution of the control variable x. The large-scale features of the field are resolved quickly in the minimization process and this corresponds to a rapid convergence at early iterations whereas the small scale features develop more slowly. Thus one has to let the minimization proceed until the convergence starts to show signs of saturation and the decrease of the cost function per iteration is very small. In this experiment, the criterion to terminate the minimization in the control assimilation is chosen to be 100 iterations and no more than 105 simulations, which follows the 4D-Var experimentation practice at ECMWF. The value of the cost function J is reduced to 0.51 compared with its initial value during this minimization, while for J_o the reduction is to 0.48.

To highlight the slow convergence during the last iterations one may note that 95% of the total reduction of J is achieved already by roughly the 35th iteration. For the quality of the following forecast, the remaining 5% is, nevertheless, important. Control assimilations with 30 iterations are also performed for 6-, 12- and 18-hour periods to get comparisons of the values of cost function and its rate of decrease.

The quasi-continuous assimilation begins with a 6-hour assimilation (A_1) where the initial point of minimization is the same background x_b as in the control assimilation. A fixed and small number of iterations (from 10 to 30) is performed with this, the shortest assimilation. This is due to the result of the barotropic experiments where no gain is achieved by saturating the minimization. The result of the minimization is then used as the initial point of the next assimilation (A_2) and the same procedure is repeated for the 12- and 18-hour periods. The termination criterion of the final 24-hour assimilation A_4 is 100 iterations or 105 simulations, as in the 24-hour control assimilation. CPU-time used for the minimization is monitored during the computations. The minimization is carried out with a quasi-Newton limited memory algorithm (M1QN3) where the diagonal elements of the Hessian are updated during the minimization (*Gilbert and Lemaréchal*, 1989). This is the aforementioned possibility to provide the preconditioning information by A_i to the subsequent assimilation A_{i+1}, i.e. the 6-hour assimilation provides preconditioning for the 12-hour assimilation and so forth. There are consequently two sets of experiments which are hereafter referred to "warm restart" and "cold restart" experiments depending on whether the Hessian information is or is not used in the assimilations A_i, respectively.

4.2 Results of the assimilation experiments

As the intention in the quasi-continuous assimilations A_1, A_2 and A_3 is to perform an approximate minimization, the number of iterations can be chosen empirically. Three alternatives with 10, 20 and 30 iterations are tested using both cold and warm restart of minimization. The number of gradient evaluations updating the Hessian is fixed to 9 in all experiments. Table 2 gives the values of J_o at the initial point of

the minimization in the assimilations A_2, A_3 and A_4 relative to the control values. Also a computing time to perform the assimilations A_1, A_2 and A_3 is given, in relative terms.

	N	$\Delta J_{o,2}$	$\Delta J_{o,3}$	$\Delta J_{o,4}$	Σt
cold restart	10	0.91	0.67	0.64	0.15
cold restart	20	0.93	0.60	0.60	0.32
cold restart	30	0.96	0.57	0.59	0.47
warm restart	10	0.91	0.67	0.63	0.15
warm restart	20	0.93	0.59	0.59	0.32
warm restart	30	0.96	0.57	0.59	0.47

Table 2. The values of J_o at the initial point of minimization in the assimilations A_2, A_3 and A_4. Here N is the number of iterations performed in the assimilations A_1, A_2 and A_3. $\Delta J_{o,2}$, $\Delta J_{o,3}$ and $\Delta J_{o,4}$ indicate the initial value of J_o in the assimilations A_2, A_3 and A_4 divided by the corresponding control value, respectively. Σt is defined as a sum of computing time of the assimilations A_1, A_2 and A_3 divided by that of the 24-hour control assimilation.

According to Table 2, J_o at the initial point of minimizations A_2, A_3 and A_4 is reduced due to performing the quasi-continuous assimilations A_i. In the 24-hour control assimilation with 100 iterations, J_o is reduced from 1 to 0.48 in relative terms. In the quasi-continuous case, J_o is already at the initial point of the minimization of the assimilation A_4 about 0.6, in relative terms. Increasing the number of iterations above 10 in the 6-hour assimilation A_1 is not beneficial for the following 12-hour assimilation A_2, whereas in the 12- and 18-hour assimilations A_2 and A_3 the reverse is true. The warm restart seems to be only slightly more efficient than the cold restart. The above reduction of the value of the cost function J_o at the initial point of minimization A_4 is achieved through a considerable computing time, from 15% to 47% of the cost of the 24-hour control assimilation (depending on N). This is compensated, however, as will be seen later.

Figure 3 shows the cost function and its gradient in the 24-hour assimilation A_4 in three cases where different number of iterations in the assimilations A_1, A_2 and A_3 is performed. Figures 3a-c show that although J_o at the initial point of minimization is considerably smaller in the 24-hour assimilation A_4 compared with the control assimilation, the subsequent rate of convergence is also notably reduced. The three J_o -curves in the Figs 3a-c saturate at about the same value, as happen also to the curves of J_b and J_c. That is quite reasonable as the observations, the formulation of variational problem and the atmospheric model are the same and therefore sooner or later the minimum of a quadratic problem has to be recovered, independently on the method.

The norm of the gradient of the cost function is shown in Figs 3d-f, which also display a running average over 10 values for each curve. The norm of the gradient is reduced by about 3 orders of magnitude in 100 iterations. A common feature in Figs 3d-f is that the norm of the gradient is initially smaller in the quasi-continuous case compared with the control assimilation and it stays smaller during the whole

minimization process. Thus, in the quasi-continuous 24-hour assimilation A_4 the latest update of the cost function is consistently closer to the minimum than in the control assimilation. Moreover, the difference is the larger the more iterations are performed in 6-, 12- and 18-hour assimilations A_1, A_2 and A_3. Figure 3d-f also shows that the use of a warm restart is beneficial for the minimization, which implies that it is possible to extract useful Hessian information even when partial sets of observations are used.

Figures 3d-f provide a way to estimate the cost reduction of a 24-hour assimilation due to the quasi-continuous approach. The norm of the gradient is a more sensitive criterion for terminating the minimization than the reduction of J itself. It takes fewer iterations in the quasi-continuous approach than in the normal one to reach any value of the norm of the gradient. An estimate of this gap is presented in Table 3. The results are based on a single case, so a general figure can be deduced only by looking at several different cases. The curve of running average over 10 gradient values of Figs 3d-f is used and any local variations on that curve have been ignored.

	N	ΔN		N	ΔN
cold restart	10	± 0	warm restart	10	15
cold restart	20	10	warm restart	20	30
cold restart	30	20	warm restart	30	40

Table 3. A subjective estimate of the cost of a 24-hour assimilation using the quasi-continuous approach. N is as in Table 2 and ΔN is defined as a difference in the number of iterations between the 24-hour quasi-continuous assimilation A_4 and the control assimilations when the same solution is reached as measured with the norm of the gradient.

When considered with the computing time (Table 2) to perform the assimilations A_1, A_2 and A_3, Table 3 indicates, that the total cost of 4D-Var can be expected to remain about the same, or slightly more expensive, with the quasi-continuous approach. However, the redistribution of the computing task implies that the cost of the final 24-hour assimilation A_4 is decreased considerably without increasing the total cost of the assimilation, provided that only a small number of iterations is performed in the assimilations A_1, A_2 and A_3. According to Fig 3d-f, the improved initial point of minimization and the warm restart contribute about equally to the cost-effectiveness of the quasi-continuous approach.

Above, only the pure computing time to run the minimization is considered. On the other hand, the model trajectory, i.e. model states over the assimilation period against which the observations are compared, grows linearly with the assimilation period as well as the amount of observations to be held in memory. If the computing time is weighted with the memory requirement, the quasi-continuous approach would seem even more beneficial compared with the normal implementation.

Figures 3a-c show that the difference in the value of J_o in different approaches may not be significant after about 60 iterations and therefore judgement on the performance of the two approaches to the assimilation problem is difficult to make. However, by concentrating on a small range of values of J_o (from $1.3 \cdot 10^5$ to $1.5 \cdot 10^5$ in Figs 3a-c) provides qualitatively similar estimates of cost reduction of the 24-hour assimilation A_4 as deduced previously from Figs 3d-f. An interesting point here is that the more horizontal the J_o -curve becomes, the larger grows the difference in the quasi-continuous and control assimilations in terms of number of iterations. The argument holds also in Figs 3d-f. Although the different approaches eventually end up to the same solution, on this limited range of iterations the difference in number of iterations seems to increase.

5. DISCUSSION AND CONCLUDING REMARKS

The concept of quasi-continuous 4D-Var has been evaluated for its effectiveness for an operational meteorological data assimilation.

In the normal implementation of 4D-Var the minimization task takes place after the current assimilation period when all the observational information is available. In the quasi-continuous approach, the assimilation is partitioned into smaller tasks. The iterative nature of the variational assimilation allows one to start the minimization with an incomplete set of observations. The practical approach envisaged here is to start the assimilation of a 24-hour period with the observations from the first 6-hour period and to perform only an approximate search of the minimum of the cost function J. As soon as more observations become available the assimilation is started again over a longer period but using the result of previous assimilation as initial point of minimization. In the present study, the length of assimilation period is gradually increased from 6 hours to 12, 18 and, finally, to 24 hours thus providing a continuously improved estimate of the analysis field. Only in the final assimilation an accurate search of the minimum of J is performed. The result of the assimilation is the same in the normal and in the quasi-continuous approach.

As discussed by *Thépaut and Courtier* (1991) and *Li et al* (1993) and (1994), the longer the assimilation period the slower the convergence. This is due to the dynamics, linear or non-linear, that modify the Hessian of the problem thus affecting the conditioning. Therefore the computing time required in the quasi-continuous approach is not as long as one would expect simply by adding together each separate assimilation period $\{T/N, 2T/N, ..., T\}$ and assuming the same convergence properties of the minimization as for the assimilation period T.

The convergence of the successive minimizations is improved in the quasi-continuous approach due to the better initial point and the better preconditioning of the minimization process. The first of these results is rather obvious as the closer to the minimum of J one begins the minimization the better. The second result

arises because it is possible to extract useful Hessian information with use of partial sets of observations even when there is a strong spatial and temporal inhomogeneity in the observational network. The improvement due to the preconditioning is of the same order of magnitude as that due to the use of a better initial point of minimization.

In the experimental assimilations with a low resolution barotropic model and using synthetic observations, the computing time of the quasi-continuous assimilation was 64% of the time for the 24-hour control assimilation. The cost of the final 24-hour quasi-continuous assimilation was 41% of the cost of the control assimilation. With a primitive equation model and with real observations the total cost of the quasi-continuous assimilation is approximately the same as that of the normal implementation. The cost of the final 24-hour quasi-continuous assimilation is in the most favourable case only 60% of the cost of the 24-hour control assimilation. The reason for this distinction in results in the cases of different models lies presumably in the simplicity of the barotropic problem and in the homogeneity of the distribution of the synthetic observations. In the quasi-continuous assimilation the trajectory of the model is shorter and there are less observations in memory until the last assimilation, resulting in a reduction in computer demand when the computing time is weighted with the memory requirement.

A possible interpretation of the cost effectiveness of the proposed approach follows. *Thépaut et al* (1993a) and (1994) showed that 4D-Var makes implicitly use of the fastest growing singular vectors to modify the covariance matrix of the background errors. The use of large scale information late in the assimilation period will modify large scales late in the assimilation together with small scales earlier in the period. The modification of the small scales early in the period can easily be below the observability. However, 4D-Var will also produce large scale changes upstream at initial time along the orthogonal of the unstable manifold. These latter changes may be observable, and therefore analysable with 3D-Var or a 6-hour 4D-Var, using observations available in the early part of the period. The interpretation is then that forcing convergence only on the large scales in the early quasi-continuous assimilations, and forcing it on all scales later in the last quasi-continuous assimilation, has analogies with multi-grid technique which has been shown to be effective for elliptic problems.

One aspect of the quasi-continuous approach is, as illustrated in Fig 1, that there appears extra time after the cut-off-time compared with the normal 4D-Var. That time originates from the faster completion of the final 24-hour quasi-continuous assimilation. Firstly, this suggests that a 24-hour 4D-Var assimilation using quasi-continuous approach can be possible even for operational centres having fairly short cut-off-time as a constraint. Secondly, there is a potential for a forecast improvement either by delaying the cut-off-time and allowing some late observations to enter the assimilation process or by searching more accurately the minimum of the cost function. The latter case causes an additional cost but the time-critical forecast task

can be started in both cases at the same wall-clock time. Computing some information about the analysis error covariance matrix to be used for cycling the 4D-Var, e.g. through a simplified Kalman filter, could be considered as well.

In the quasi-continuous approach the assimilation code remains unchanged whereas the organisation of the computing resources will become more complicated. The demand for system time increases as the initialization of assimilation tasks as well as the result of minimization and the preconditioning information has to be handled several times. From a pure operational point of view, the quasi-continuous approach may be tolerable as the 4D-Var assimilation task is divided into smaller parts of which only the last is time and memory critical. During the period T, the scheduling of the quasi-continuous assimilations can be flexible. The observation handling has to be partly reorganized as there will be N different assimilations and preprocessing tasks to be executed and some extra checking may become necessary. Although there will appear some extra work in ordering and checking of observations, the irregular arrival of observations does not affect the result of assimilation.

A less trivial aspect which needs to be studied concerns applying non-Gaussian error statistics to the observations. Error statistics of this kind may be used in quality-control of observations which is integrated into the variational assimilation giving smooth transition from datum rejection to its acceptance rather than strict limits applied as a separate step independently on the assimilation (*Lorenc and Hammon*, 1988). The topology of the functional J may in this case become complicated and eventually possess several minima. In such a case, there is a risk during the course of minimization to fall into a non-optimal secondary minimum. It may be therefore necessary not to saturate the minimization on the early quasi-continuous assimilations and to ensure the following assimilations to proceed to an optimal solution.

In the near future, the next generation remote sensing instruments will provide a considerably larger amount of atmospheric measurements than at present. The observations with high spatial and temporal resolution eventually leads to the necessity of more continuous data processing and assimilation methods. The quasi-continuous approach is one possible step in this direction utilizing the possibilities provided by the variational data assimilation.

ACKNOWLEDGEMENTS

We are grateful to our colleagues at ECMWF for many discussions and helpful comments, particularly Florence Rabier, Per Undén and Erik Andersson who are warmly acknowledged. Anthony Hollingsworth made a very welcome contribution by reading the manuscript and by providing a possible explanation for the cost effectiveness of the proposed approach. Thanks are also due to INRIA (Institut National de

Recherche en Informatique et en Automatique), Le Chesnay, France, for providing the minimization algorithm (M1QN3).

REFERENCES

Courtier, P, 1987: Application du contrôle optimal à la prévision numérique en Météorologie. Thèse de doctorat de l'université, Paris VI.

Courtier, P and O Talagrand, 1990: Variational assimilation of meteorological observations with direct and adjoint shallow-water equations. Tellus, **42A**, 531-549.

Courtier, P, E Andersson, W Heckley, G Kelly, J Pailleux, F Rabier, J-N Thépaut, P Undén, D Vasiljevic, C Cardinali, J Eyre, M Hamrud, J Haseler, A Hollingsworth, A McNally and A Stoffelen, 1993: Variational assimilation at ECMWF. ECMWF Technical Memorandum No.194, 84pp.

Courtier, P, J-N Thépaut and A Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. Q J R Meteor Soc, **120**, 1367-1387.

Derber, J C, D F Parrish and J G Sela, 1992: The SSI analysis system and extensions to 4D. ECMWF Workshop Proceedings on Variational Assimilation With Special Emphasis on Three-dimensional Aspects, ECMWF, 9-12 November 1992, 15-35.

Gandin, L S, 1963: Objective analysis of meteorological fields. Translated from the Russian by the Israeli Program for Scientific Translations (1965).

Gilbert, J C and C Lemaréchal, 1989: Some numerical experiments with variable storage quasi-Newton algorithms. Math Prog, **B25**, 407-435.

Heckley, W A, P Courtier and J Pailleux, 1992: The ECMWF variational analysis: general formulation and use of background information. ECMWF Workshop Proceedings on Variational Assimilation With Special Emphasis on Three-dimensional Aspects, ECMWF, 9-12 November 1992, 49-94.

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

Järvinen, H, 1993: A direct conversion of model algorithms to their adjoints exemplified with a barotropic atmospheric model. Dept of Meteor, Univ of Helsinki, Report No 42, 41pp.

Li, Y, I M Navon, P Courtier and P Gauthier, 1993: Variational data assimilation with a semi-lagrangian semi-implicit global shallow-water equation model and its adjoint. Mon Wea Rev, **121**, 1759- 1769.

Li, Y, I M Navon, W Yang, X Zou, J R Bates, S Moorthi and R W Higgins, 1994: Four-dimensional variational data assimilation experiments with a multilevel semi-lagrangian semi-implicit general circulation model. Mon Wea Rev, **122**, 966-983.

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

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

Lorenc, A C and O Hammon, 1988: Objective quality control of observations using Bayesian methods. Theory and practical implementation. Q J R Meteorol Soc, **114**, 515-543.

Pailleux, J, 1989: Design of a variational analysis: organization and main scientific points. Computation of the distance to the observations. ECMWF Tech Mem, **150**, 23pp.

Pailleux, J, W Heckley, D Vasiljevic, J-N Thépaut, F Rabier, C Cardinali and E Andersson, 1991: Development of a variational assimilation system. ECMWF Tech Mem No: 179, 51pp.

Pailleux, J, 1993: Organisation of 3D variational analysis within the "IFS/Arpège" project. Future plan at Météo France. ECMWF Workshop Proceedings on Variational Assimilation With Special Emphasis on Three-dimensional Aspects, ECMWF, 9-12 November 1992, 37-47.

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

Rabier, F, P Courtier, J Pailleux, O Talagrand and D Vasiljević, 1993: A comparison between four-dimensional variational assimilation and simplified sequential assimilation relying on three-dimensional variational analysis. Q J R Meteorol Soc, **119**, 845-880.

Rinne, J and H Järvinen, 1993: Estimation of the Cressman term for a barotropic model through optimization with use of the adjoint model. Mon Wea Rev, **121**, 825-833.

Shaw, D B, P Lönnberg, A Hollingsworth and P Undén, 1987: Data assimilation: The 1984/85 revisions of the ECMWF assimilation system. Q J R Meteorol Soc, **113**, 533-566.

Talagrand, O, 1991: The use of adjoint equations in numerical modeling of the atmospheric circulation. Proceedings of the first SIAM Workshop on Automatic Differentiation. Breckenridge, Colorado, January 6-8, 1991, 169-180.

Thépaut, J-N and P Moll, 1990: Variational inversion of simulated TOVS radiances using the adjoint technique. Q J R Meteor Soc, **116**, 1425-1448.

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

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

Thépaut, J-N, D Vasiljevic, P Courtier and J Pailleux, 1993b: Variational assimilation of conventional meteorological observations with a multilevel primitive-equation model. Q J R Meteorol Soc, **119**, 153-186.

Thépaut, J-N, P Courtier, G Belaud and G Lemaître, 1994: Dynamical structure functions in a four-dimensional variational assimilation: a case study. To be submitted to Q J R Meteorol Soc.

Vasiljević, D, C Cardinali and P Undén, 1992: ECMWF 3D-variational data assimilation of conventional observations. ECMWF Workshop Proceedings on Variational Assimilation With Special Emphasis on three-dimensional aspects. 9-12 November 1992, 389-426.

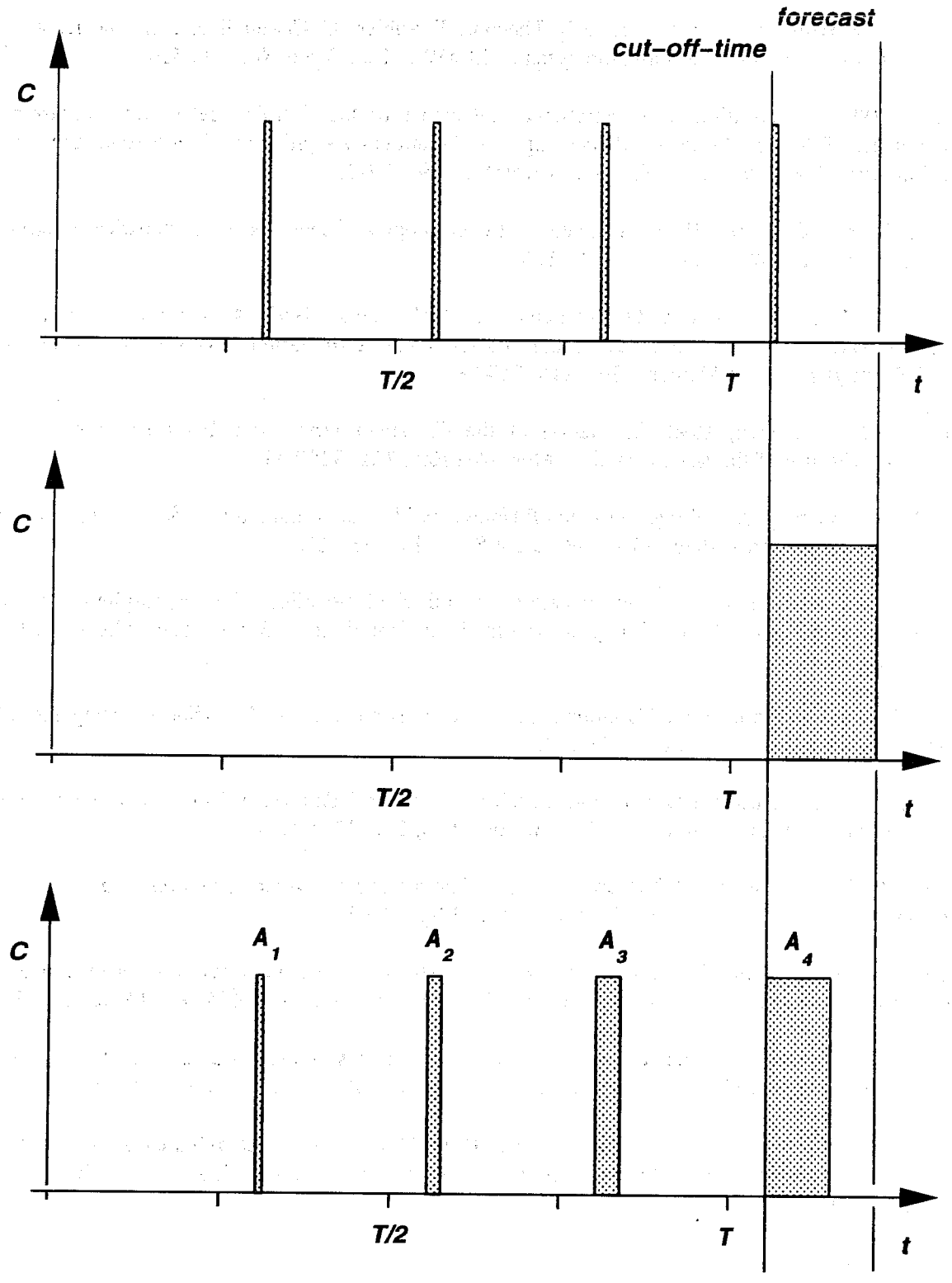

Fig 1 A schematic presentation of an OI implementation (uppermost panel), of a normal 4D-Var (middle panel) and that of a quasi-continuous 4D-Var approach (lowermost panel). The ordinate ***C*** describes computing power demand due to the assimilation of observations. Vertical intersections denote two time-levels: the cut-off-time of normal 4D-Var and the starting of the time critical forecast run. A_1, A_2, A_3 and A_4 refer to the quasi-continuous assimilations.

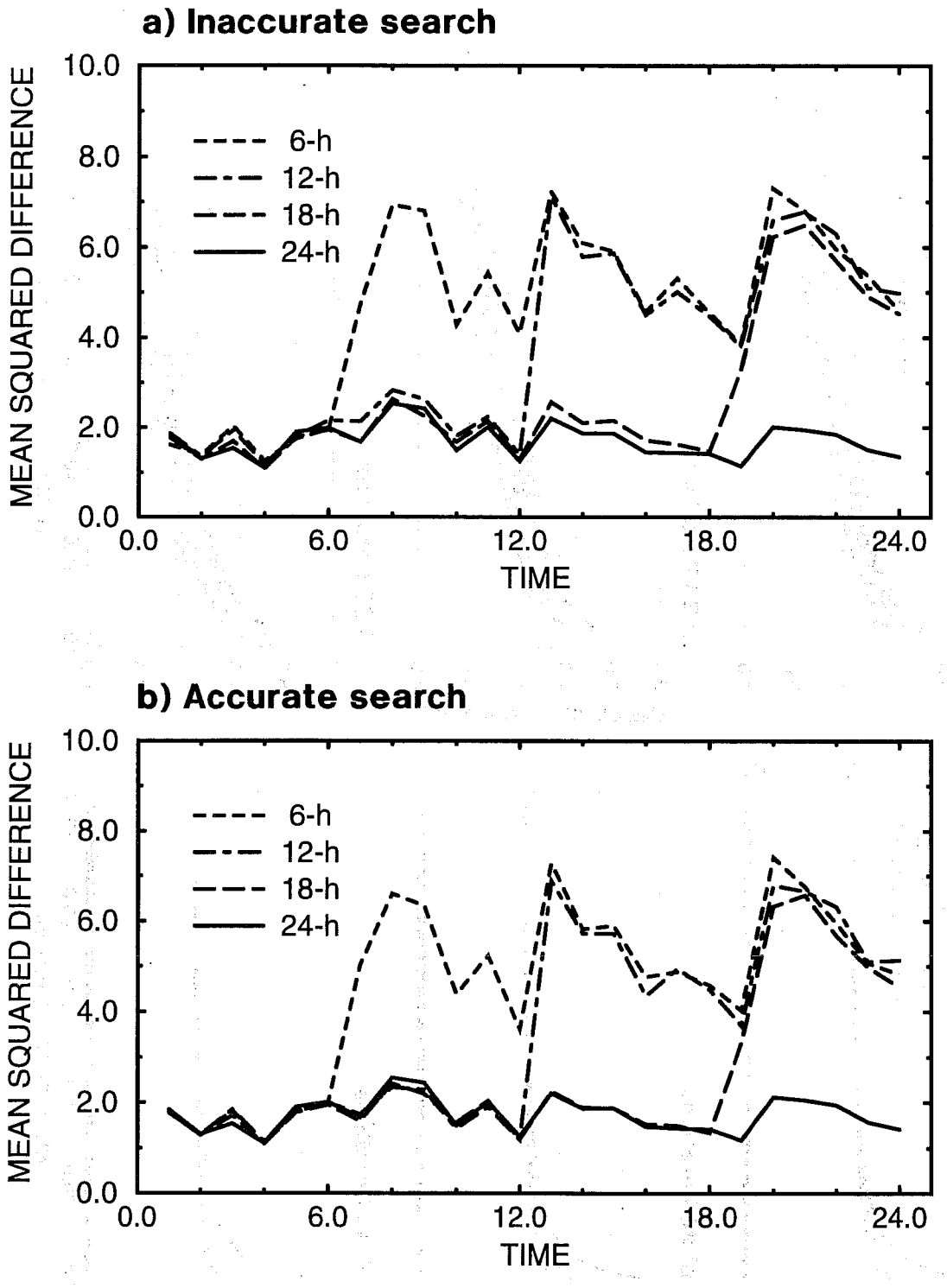

Fig 2 Mean squared difference between the model state and the observations over the 24-hour assimilation period in the quasi-continuous assimilations.
(a) The case where an inaccurate search of minimum of cost function is performed in assimilations A_1, A_2 and A_3.
(b) The same, but for an accurate search. Note that the solid line is virtually the same in (a) and (b).

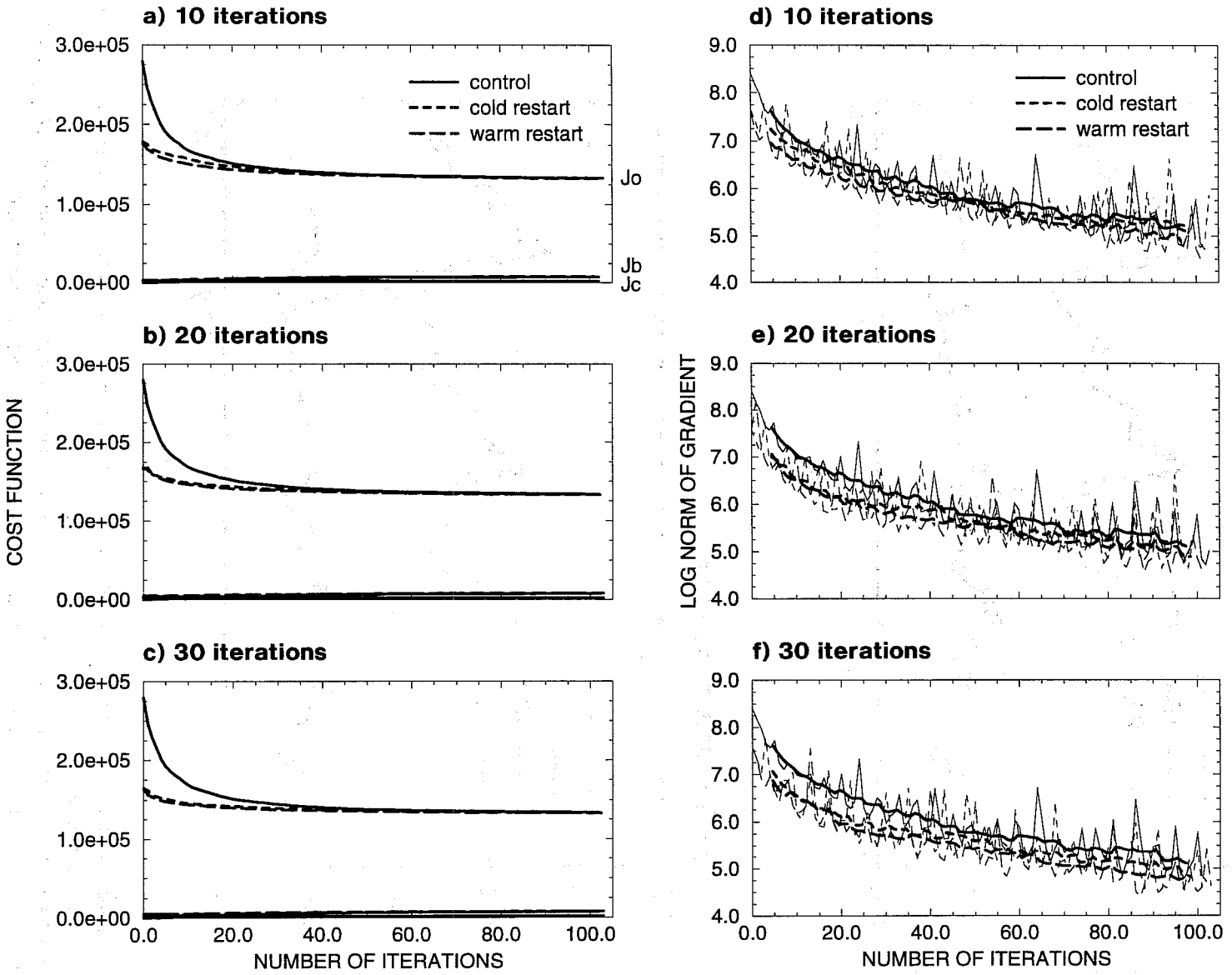

Fig 3 The reduction of terms J_b, J_0, J_c of the cost function (a, b, c) and norm of the gradient on logarithmic scale (d, e, f) in 24-hour assimilations in the cases where 10 (a, d), 20 (b, e) or 30 (c, f) iterations are performed in the assimilations A_1, A_2 and A_3. Solid line is for control, dashed for cold restart quasi-continuous and long dashed for warm restart quasi-continuous assimilations, respectively. In (d, e, f) also a running average over 10 gradient values is plotted (thicker lines).