179

Development of a variational assimilation system

J. Pailleux, W.A. Heckley, D. Vasiljević, J N. Thépaut, F. Rabier, C. Cardinali and E. Andersson

Research Department

August 1991

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

Development of a Variational Assimilation System

Document presented to the 19th Session of the Scientific Advisory Committee, 1991.

1.   INTRODUCTION

The impact of TOVS data on the analysis and forecast has always been large and positive in the Southern Hemisphere. However, the results in the Northern Hemisphere have never shown any clear positive impact of TOVS data. Since 1979, many impact studies have been carried out to assess the impact of TOVS data on analyses and forecasts, (*Halem et al.* (1982); *Gilchrist* (1982); and *Gallimore and Johnson* (1986)). At ECMWF, a comprehensive set of experiments on the quality and impact of TOVS data is documented in *Andersson et al.* (1991) and *Kelly et al.* (1991). The impact of TOVS in the Northern Hemisphere is neutral on average, but some cases of positive impact are found, as well as some cases of negative impact which can often be traced back to "misuse" of the satellite data: (*Kelly et al.* (1991), *Kelly* (1990)). This is somewhat disappointing, taking into account the volume of TOVS data and the potential information which it contains. The clear signal from most of these impact studies is that the TOVS are not used in an optimal way in most operational analysis schemes. The reasons for this "sub-optimality" are, to a large extent, understood. Generally the satellite information is passed to the analysis in the form of retrieved profiles (usually called SATEMs) which are only a crude representation of the genuine satellite information and also contain rather inaccurate auxiliary information (such as statistical information or an initial profile which often has a strong component of climatology), *Pailleux* (1988). Then, the Optimal Interpolation analysis using these retrieved profiles, cannot extract the radiance information correctly. A few attempts have been made to use the radiances directly in Optimal Interpolation (see *Durand, 1985*); these attempts were limited by the necessity of an assumption of linearity between radiances and temperatures.

Since May 1991 no TOVS data have been used in the tropics or in the Northern Hemisphere troposphere within the ECMWF operational analysis.

A linear Optimal Interpolation analysis cannot extract properly from the observations any quantity non-linearly related to the forecast model variables. This is a strong limitation, both for currently available data, such as TOVS, and also for many future observing systems, such as scatterometer observations. Not even all current conventional data can be used by OI; for example, an observed wind speed cannot be used on its

own, without an observed wind direction. The natural tool to cope with this non-linearity is a variational analysis based on the notion of an adjoint operator (see *Pailleux*, 1988).

The direct use of radiances in a three-dimensional variational analysis, together with all the other observation types, through a direct radiative transfer model and its adjoint, is a natural and attractive solution. Another approach, which is also under test at ECMWF, consists of using the 6 hour forecast to perform a variational retrieval in the vertical (1D-VAR), before using these retrievals in the Optimal Interpolation analysis.

Historically, the adjoint operators were first studied in meteorology in the context of 4 dimensional assimilation (rather than three-dimensional analysis): the adjoint of a forecast model is the appropriate tool to fit a model "trajectory" to observations over a time period (t_o, t_1), and so provide a "clean" solution to the general four-dimensional continuous assimilation problem. These four-dimensional aspects are documented in *Lewis and Derber* (1985), *Le Dimet and Talagrand* (1986), *Talagrand and Courtier* (1987), *Courtier and Talagrand* (1987). Compared with a 6-hour intermittent assimilation system (such as that currently operational at ECMWF and in many other forecasting centres), a four-dimensional continuous system presents many potential advantages: observations may be used in a way fully consistent with the model equations, full use can be made of observations reporting with a high time frequency, additional possibilities exist to tackle "spin-up" problems. The ability to investigate these aspects through four-dimensional experimentation is a strong motivation for developing a global variational assimilation scheme able to work both in a three-dimensional and a four-dimensional mode. A four-dimensional assimilation scheme is likely to be expensive in computer resources, as it has to iterate a calculation which includes a direct integration of the forecast model from time t_o to t_1, and a backward integration of the adjoint model from t_1 to t_o. In addition it requires the storage of the model trajectory. Experimentation is required to answer the question whether operational assimilation can be done in a four-dimensional mode. A key issue is to evaluate, in a quantitative way, to what extent the model dynamics can help the assimilation by, for example, reconstructing some typical features of the atmosphere.

The Optimal Interpolation scheme cannot work operationally without a data selection algorithm. Two data selection strategies have been used. The point data selection strategy chooses a limited number of observed data around each grid-point and for each analysis level. In the box method used in the ECMWF system (*Lorenc*, 1981), an observation set is selected to analyze a whole volume through one Optimal Interpolation linear system of equations. In any case the data selection algorithm generates some noise in the final analysis because different observations are used from one point to another to correct the first-guess. Although we do not have any quantitative assessment of the degradation coming from the data selection scheme, a global

2

analysis using all the data in one calculation is attractive because it would remove this deficiency and, at the same time, simplify the analysis code. It would perhaps also give an opportunity to assess the impact of data selection.

The need to study these three-dimensional and four-dimensional assimilation issues in a consistent way (together with other requirements, e.g. the use of a tangent linear model to investigate the time evolution of errors), led ECMWF and METEO-FRANCE to develop, in common, a new system called Integrated Forecasting System (IFS) at ECMWF and "ARPEGE" in France. The main features of IFS/ARPEGE are described in section 2. They define the framework in which the variational assimilation work has been carried out. In section 3, the different variational assimilation developments which have been made in the IFS are documented. The results of the different experiments which have been carried out are given in sections 4 and 5: section 4 for three-dimensional experiments, section 5 for four-dimensional experiments. Some remarks about the different minimization algorithms which have been used for this work are given in section 6. Section 7 concludes the paper and describes the prospects and plans at ECMWF in the area of variational assimilation. We also indicate other applications of the IFS system.

2. GENERAL FEATURES OF THE "INTEGRATED FORECASTING SYSTEM" (IFS)

The development of the IFS/ARPEGE system began at the end of 1987. It was designed from the beginning to provide all the modules needed to perform three-dimensional and four-dimensional variational assimilation with real observations and also with simulated ones ("observed" fields either in grid points or spectral space):

- A forecast model and its adjoint (for four-dimensional assimilation).
- The observation operators needed to compare the model with observations and to compute an observation cost function (J_o: metric of model minus observations distance) and its gradient.
- The first guess operator needed to compare the model with a guess and to compute a guess cost function (J_g: metric of model minus guess distance) and its gradient.
- Mass/wind balance operators.
- General minimization algorithms.

Talagrand (1988), *Pailleux* (1988) and *Pailleux* (1990) discuss the role of these modules in a variational assimilation.

The IFS/ARPEGE system includes a normal operational forecasting model. It is planned to have a version of the ARPEGE model operational in France by the end of 1991 and for the IFS model to be operational at ECMWF in 1992. The main forecast model of the IFS/ARPEGE system is a multilevel primitive equation

spectral model with the option of a variable resolution in the horizontal (*Courtier and Geleyn*, 1988), and a hybrid coordinate in the vertical (*Simmons and Burridge*, 1981). Options have been prepared for:

- Semi-Lagrangian scheme.
- Reduction of the number of grid points near the poles on the collocation grid.
- Rotation of the pole (this option, combined with a stretching factor making the horizontal resolution higher near one chosen point, may enable the system to provide the functions of current limited area models).

An Optimum Interpolation analysis is also included in the IFS/ARPEGE system in order to meet the French operational requirements.

The IFS/ARPEGE system also incorporates several research related tools: -

- Vorticity equation and shallow water equation models.
- Tangent linear models which are useful to study the time evolution of model errors (though they are not needed as such for four-dimensional assimilation experiments).
- Kalman filtering (also a key technique for assimilation, not described in this paper).
- Validation tools for testing the correctness of an adjoint code or the correctness of a gradient.

The code has been developed in such a way that the various applications can be used in combination with the various models by turning on or off appropriate switches. The dynamical part of the forecasting models was developed first, together with their corresponding tangent linear and adjoint versions: (1987-1988). The strategy which has been followed and the tools which have been used for validating these developments are described in *Thépaut and Courtier* (1991). The code development work has continued in several parallel streams: -

- Rotation of the pole, normal mode initialization, reduced grid near the poles.
- Code for Semi-lagrangian advection (mainly done in France).
- Incorporation of ECMWF and DMN physical packages.
- Development of the variational analysis code (mainly done at ECMWF).
- Development of an Optimum Interpolation scheme (in France; ECMWF has no plans to change its Optimal Interpolation code).

The last two modules have been developed with common observation operator routines: for both variational analysis and Optimal Interpolation one needs to interpolate the model both in the horizontal and the vertical in order to compare it with the different observed quantities. The model fields used for this comparison are either the control variable (updated at each iteration of the variational analysis) or the guess (for Optimal

Interpolation). This follows our general policy to avoid coding the same computation or the same operator twice, thus improving the internal consistency of the IFS/ARPEGE system. The bulk of the analysis code development took place in 1989-90. Since the beginning of 1991, the variational analysis has entered the debugging and testing phase.

The first goal of the analysis code tests was to make sure that each different contribution to the cost function of the variational analysis had a correct gradient, including the contributions of different model variables to J_g and the contributions of the different observation types and different variables to J_o. The standard validation tool is the gradient test option (*Thépaut and Courtier*, 1991). Having developed the code for computing a given contribution $J(X)$ to the variational analysis cost function and its gradient ∇J, one can test that the vector ∇J is the correct gradient of the function $J(X)$ with respect to the control variable through a check based on the Taylor formula:

$$t = \lim_{\delta X \to 0} \frac{J(X + \delta X) - J(X)}{< \nabla J, \delta X>} = 1$$

δX is a perturbation to the control variable X. In the standard gradient test algorithm, δX is chosen proportional to the gradient: $\delta X = -\alpha \nabla J$, as this choice normally leads to a reasonable scaling of the various components. This choice is arbitrary and, in some circumstances, it was found more convenient to choose δX proportional to X (i.e. $\delta X = \alpha X$). The gradient test checks that the above ratio tends to 1, linearly in α, over a wide range of magnitude of α. The example shown in Table 1 evaluates the gradient of the TOVS radiance cost function from 1200 simulated observations with respect to all the model variables in spectral space. The linearity can be visually checked, as shown in table 1, by the number of decimals equal to 9 or to zero in the value of the above ratio, printed for different values of α increasing by a factor 10 from one line to the next. We have found that the test is a powerful debugging tool. Most of the time, a lack of linearity in the gradient test was due to small bugs. However, in a few tests related to some observation operators, this lack of linearity was found to be due to a "small amount of non-differentiability" in the operators.

α	t
1.00 E-13	1.3545476596041
1.00 E-14	1.0267112532746
1.00 E-15	1.0026520613185
1.00 E-16	1.0002650049736
1.00 E-17	1.0000258019939

Table 1:   Gradient test function for different values of α which specifies the size of the perturbation. The cost function being tested corresponds to the contribution of 1200 simulated TOVS radiances, and we are checking its gradient with respect to all the model variables in spectral space.

3.   VARIATIONAL ANALYSIS IN THE IFS

In the global data assimilation problem one has to find a model trajectory which fits "reasonably" the observations available on an assimilation period (t_o, t_1), and which also fits "reasonably" the most recent forecast valid for t_o (first guess). This trajectory is entirely determined by the vector X containing the ensemble of model variables at time t_o: once the required X has been computed, the required trajectory will be obtained by an integration of the forecast model from "state X" at time t_o to time t_1. X may be taken as <u>control variable</u> of the following global minimization problem: minimize the <u>cost function</u> $J(X) = J_g + J_o$, where

J_g is defined as the distance from X to the first guess X_g;

J_o is defined as the distance from X to the observations:

In the four-dimensional problem the model has to be integrated from t_o to the appropriate observation time in order to compare X with the observation, then the adjoint model has to be integrated back to t_o in order to obtain the appropriate gradient with respect to X.

A three-dimensional variational analysis can be designed like a general four-dimensional system in which the model integrations are switched off: one then compares X directly with the observations made at time t_o (or around t_o). It can be shown (*Lorenc*, 1988), that a global Optimal Interpolation is equivalent to the minimization of the cost function:

$$J(X) = J_g + J_o$$

with $J_g = (X-X_g)^t\, P^{-1}\, (X-X_g)$

$J_o = (KX-d)^t\, O^{-1}(KX-d)$

where X_g is the first-guess, P the covariance matrix of first-guess errors, d is the vector containing all the observed data, K is the product of the operators transforming the control variable X into the equivalent of each observed quantity, O is the covariance matrix of observation errors (which also contains the representativeness error).

Minimization of $J(X)$ with respect to X requires the following steps:

a)   Provide an initial estimate for X.

b)   Compute $J(X)$ and its gradient with respect to X.

c)   Pass X and $J(X)$ to a minimization scheme which computes a more accurate estimate of X.

d)   Iterate on b) and c) until a reasonable convergence is achieved.

In general the cost function may have any number of terms -

$$J = J_g + J_o + J_c + ...$$

where J_c, for example, may contain balance constraints. The use of a J_c is one technique to ensure a proper mass wind balance.

The variational analysis part of the IFS/ARPEGE system reduces to the computations of J_o, J_g, J_c and their gradients. The gradient of J is given by

$$J_X = \nabla J = \nabla_X (J_g + J_o + J_c + ...) = 2\,(P^{-1}\, X + K^t\, O^{-1}\,(KX-d)) + \nabla J_c + ...,$$

and the minimizing solution is found by setting this expression to zero. For example, considering only J_g and J_o

$$\tfrac{1}{2}\, J_X = P^{-1}X + K^t O^{-1}(KX - d) = 0.$$

In methods of steepest descent the search direction is the direction of the local gradient J_X. This can be inefficient in the region of steep valleys in the cost function. A better approach is the quasi-Newton, or

7

variable metric methods. In such an approach the local gradient is pre-multiplied by J_{XX}^{-1}, the inverse of the Hessian matrix of J. If J is a strictly concave quadratic function then pre-conditioning in this way ensures the minimum will be found in a single step. In other cases it is still a great improvement on steepest descent. In this approach one effectively takes a second order rather than first order local approximation to J. In the case when K is a linear operator (independent of X) J_{XX} is given by

$$\tfrac{1}{2} J_X = P^{-1} + K'O^{-1}K.$$

For a linear problem this will result in optimal pre-conditioning of the minimization. However the inverse of the Hessian is difficult to compute and store, involving as it does all the observational operators as well as the guess operator. As a starting point it will be assumed that the observation error operator has the same form as P^{-1}. This may not be too bad an assumption as the prediction errors in an assimilation will be closely related to the analysis errors. With this assumption the Hessian is proportional to P^{-1}.

P^{-1} is a square matrix with dimension given by the number of degrees of freedom of the forecast model. In the current code P^{-1} only includes the model state variables and is block diagonal. Thus forecast errors are assumed uncorrelated between variables in spectral space, which corresponds to an isotropy assumption on the sphere. Further, separability is assumed between the 'vertical' and 'horizontal' components of P^{-1}. Even so, for one of these block components, such as the vorticity field at a particular model level, the size of the block matrix is far too large to store. Thus, J_{XX} as a whole cannot be passed to the minimization algorithm. At the moment only the diagonal component of the Hessian is provided to the minimization. This means that the minimization will perform best when P^{-1} is nearly diagonal.

3.1 Guess Cost Function

The guess cost function is given by

$$J_g = (X - X_g)^t P^{-1} (X - X_g)$$

where, as mentioned above, separability is assumed between the 'horizontal' and 'vertical' components of P: -

$$J_g = (X - X_g)^t V^{-1} H^{-1} (X - X_g).$$

The formulation of H^{-1} and V^{-1} is described in Appendix A. For the moment, note that H contains within it the covariance matrix in spectral space of the first-guess errors within model levels; V contains the

correlation matrix of first-guess errors between model levels, and the latter is assumed to have no geographical variability. All of the spatial variability, both horizontal and vertical is contained within H.

J_g may be reformulated in terms of the normalized departures of the models variables from the first guess -

$$J_g = \left(\frac{X - X_g}{\sigma}\right)^t V^{-1} h^{-1} \left(\frac{X - X_g}{\sigma}\right)$$

where the σ's describe the spatial variability of the forecast errors (see Appendix B). A diagonal form may be assumed for h^{-1} (which is now a correlation matrix), but not for V^{-1}. The form of h^{-1} is chosen such as to give isotropy and homogeneity of the correlation function on the sphere.

Since the forecast errors, σ, are defined in physical space and the cost function is calculated in spectral space the division by forecast errors involves a sequence of operations -

1)   Difference X and X_g in spectral space

2)   Convert from vorticity, divergence to winds D^{-1}

3)   Transform to grid-point space S^{-1}

4)   Normalize with respect to forecast errors σ N

5)   Transform to spectral space S

6)   Convert from winds to vorticity, divergence D

$$x = D\ S\ N\ S^{-1}\ D^{-1}\ (X - X_g)$$

$$J_g = a\ x^t\ V^{-1}\ h^{-1}\ x$$

a is equal to 1 except for vorticity, divergence, when it takes on the form $(-1/C_n)$, where C_n are the eigenvalues of the Laplacian operator on the sphere. This uses the Green's formula:-

$$K(n) = -\frac{1}{C_n} \underset{n,m}{\Sigma} (\ |\zeta_n|^2 + |\delta_n|^2\ ).$$

This ensures that the normalization which is applied to the winds in physical space results in suitably normalized vorticity and divergence fields in spectral space. The gradient with respect to x is given by: -

$$\nabla_x J_g = a\ V^{-1}\ h^{-1}\ x.$$

In order to obtain the gradient with respect to X the adjoints of the above operations have to be applied in reverse sequence.

$$\nabla_X J_g = (D^{-1})^*(S^{-1})^* N^* S^* D^* \nabla_x J_g$$

Where * denotes an adjoint operator.

In general the operator N cannot be considered a quadratic term and the Gaussian grid is insufficient to avoid aliasing. However, as described in Appendix B, care is taken in the specification of the σ values so as to ensure that aliasing errors are reduced to acceptable levels.

3.2 Control of noise, and balance constraints

In the assimilation context, one is interested in finding an analyzed state which is close both to the observations and to the slow manifold. Within the variational approach it is possible to include balance constraints in a number of different ways.

Two approaches have so far been considered :

i)   The cost functions may be formulated in terms of initialized model fields, $NMI(X)$, so that

$$J_g = (NMI(X) - X_g)^t \, P^{-1} \, (NMI(X) - X_g)$$

$$J_o = (K(NMI(X)) - d)^t \, O^{-1} \, (K(NMI(X)) - d)$$

This involves a change of variable by performing a non-linear normal mode initialisation (NMI) on X, the adjoint of which is needed in calculating the gradient of the cost function with respect to X. Implementation is straightforward as the NMI and its adjoint are simply operators which are applied at the appropriate points in the chain.

ii) The second approach consists of introducing a cost function J_c which contains a penalty term on the tendency of gravity modes G

$$J_c = \alpha_c |dG/dt|^2$$

This is carried out by computing the tendency of the gravity modes of the analyzed state through one time-step of the model; and the adjoint through one time step of the corresponding adjoint model.

It has been found in *Courtier and Talagrand* (1990) and *Thépaut and Courtier* (1991) that the first approach acts to speed the convergence (since the minimization is performed in a reduced space). However, if the number of iterations increases, the NMI process is inverted by the minimization and the minimizing solution contains gravity components. This is discussed further in section 5.1.

10

The second solution is the only one which really constrains the solution to be free of gravity waves, but the convergence turns out to be slow. A combination of the two approaches has been found beneficial in a four-dimensional context. Both approaches have been coded.

A somewhat different solution has been adopted in the variational analysis which became operational in June 1991 in the National Meteorological Center of Washington (*Parrish and Derber*, 1991). The model variables are constrained in order to stay close to the equilibrium of the linear balance equation applied on the model levels. This is achieved by a choice of control variable which takes into account the balance equation. The NMC analysis variables are:

- Departures to the 6 h forecast for vorticity and divergence;
- Departures to the balance equation solution of the temperature departures to the 6 h forecast.

By assigning appropriate statistics to the errors on these variables, a balance is achieved which is good enough to obviate the need for normal mode initialization. The variational analysis of specific humidity is performed separately.

3.3   Choice of control variable and the influence of balance constraints.

Let us first consider the simple case with no vertical coupling: V^{-1} is a unit matrix. In this case J_g has the form

$$J_g = \left(\frac{X - X_g}{\sigma} \right)^t h^{-1} \left(\frac{X - X_g}{\sigma} \right)$$

where h^{-1} is diagonal.

Since the $\sigma's$ are spatially varying the P matrix will be far from diagonal, and the problem is not well conditioned if a diagonal form for the Hessian is supplied to the minimization. If the σ values are taken as constant over η levels, P is diagonal, and exact minimization of J_g alone can be accomplished in 1 iteration.

If, however, the control variable is re-defined as

$$x = \left(\frac{X - X_g}{\sigma} \right)$$

(or as $x = X/\sigma$), then the Hessian is simply h^{-1}, which is diagonal. Such a change of control variable improves the pre-conditioning. Indeed, in a two-dimensional calculation (where V^{-1} is a unit matrix) the exact solution for J_g alone is found in a single step, even with full geographical variability of the forecast errors.

The addition of balance constraints as outlined in section 3.2 complicates the minimization issue. For the most efficient minimization one should include the Hessian of the NMI (and of J_c, which will be similar) in the pre-conditioning. No attempt to do this has so far been made.

3.4   Observation Operators

The general strategy for setting up the observation operators has been outlined by *Pailleux* (1989) which is the design document for the use of observations in the variational analysis. The plan is to first develop the operators needed to use the data which are already used in the ECMWF operational Optimal Interpolation analysis, in a way which reproduces the Optimal Interpolation assumptions as far as possible. Then, in a second step, these operators will be varied to do research on the use of observations; and new operators will be developed for new data as required. An exception has been made to this strategy for TOVS, as will be explained later.

The operators which are common to all observations are:
*   The inverse spectral transforms from spectral fields to grid point representation.
*   the horizontal interpolation from the grid points to the observation points; this is always applied before the vertical part of the operators, consequently the horizontal interpolation is performed directly on the model variables in the hybrid coordinate. The current interpolation scheme is bi-linear interpolation in latitude/longitude from the four nearest grid points.

3.4.1   Conventional Observation Operators

All the operators corresponding to observations which are used in the ECMWF operational Optimal Interpolation analysis, with the exception of 10 m wind and 2 m relative humidity data from SYNOPs, SHIPs and buoys, have been developed in the IFS/ARPEGE system, as well as their adjoints which are needed for the gradient computations. To be precise, the following observations are used at present:

*   From surface observations (SYNOPs, SHIPs, DRIBUs), surface pressure data is used at the station level or reduced to mean sea level, or geopotential.

- From radiosondes (TEMPs), the data used are geopotential heights, wind components and relative humidities. Temperature operators have been developed as well, but no attempt has yet been made to use all the significant level data.
- From PILOTs, wind components are used.
- From AIREPs, wind components are used (not temperatures).
- From SATOBs, wind components are used.

The vertical part of a typical observation operator K usually consists of vertical interpolations from the hybrid model levels to the observation level. The vertical interpolation is performed using the standard postprocessing routines whose adjoint and tangent linear routines have been developed. The version of the routines which have been developed is currently based on the simplest assumption: all the variables are interpolated linearly in p between two hybrid levels. Special treatments are made at the bottom and the top of the atmosphere. The development of the variational analysis was actually the first opportunity to validate the adjoint vertical postprocessing routines in the context of a real size application. Having only one set of vertical interpolation routines (+ adjoint + tangent linear), and using them for all applications (variational analysis, model postprocessing, model dynamics), is consistent with the policy that only one routine is used to perform one particular function in the IFS.

Since it is reasonable to assume that the observation errors are not correlated between observation types, J_o can be split into different contributions according to each observation type:

$$J_o = J_{synop} + J_{RS} + J_{sat} + \ldots$$

Moreover, for most of the observation types, the errors can be assumed uncorrelated in the horizontal (e.g. it is natural and common practice in Optimal Interpolation to assume that the observation errors of two different radiosondes are independent). Most of the time J_o is then the sum of the different contributions from each individual observation. These contributions can be computed independently: the corresponding quadratic forms are still in the form $(KX-d)^t\ O^{-1}\ (KX-d)$, but the size of matrix O is very small (it is only of dimension 1 when there is only one observed datum). For a single radiosonde which observes height Z, the wind W and humidity RH, the cost function can also be split into $J_z + J_w + J_{RH}$.

In order to make the computation efficient on a vector computer with several processors, the vertical part of the observation operators is not performed observation by observation. Observations of the same type are first collected in "packets" or "sets". Each observation set can contain up to 200 observations of the same type (e.g., 200 radiosondes, 200 SYNOPs). Then different sets can be treated in parallel on different

processors. On each processor, the vertical processing is normally vectorized, the size of the vectors being the size of the observation sets.

The conventional observations entering the variational analysis have previously gone through all the operational quality control checks. The variational analysis experiments are thus performed with the same observed data as the operational ECMWF analysis, which makes the use of the Optimal Interpolation analysis as a reference easier. Also all the quality, and other "event" flags are passed to the variational analysis, which makes the final choice of data in the variational analysis fully flexible.

3.4.2 TOVS Operators

Ideally, one should use the TOVS raw radiances directly in the variational analysis. However, accurate computation of the model equivalent of raw radiances (from X) requires a good description of the clouds (which cannot be provided at present). For this reason we will instead use the cloud-cleared radiances instead of the raw radiances in the variational analysis. This is a realistic operational scenario, as the cloud-cleared radiances produced operationally in NESDIS (Washington) are available to ECMWF (together with the NESDIS retrieved profiles) with a 120 km resolution in the horizontal. Because the cloud-cleared radiances do not undergo the Optimal Interpolation quality control, a special quality control algorithm has been developed (called PRESAT). PRESAT has been in operation at ECMWF since May 1991 and the cloud-cleared radiances will be checked by this algorithm before entering the variational analysis.

To compute the distance of the model to the radiances, we need the inverse spectral transforms and the horizontal interpolation, as mentioned in the previous section. We also need a radiative transfer model T_r to compute the model's equivalent of TOVS radiances from a model temperature/humidity profile at each TOVS observation point. T_r must be preceded by a vertical interpolation from the model levels to the pressure levels which are required by the radiative transfer model. The adjoint T_r^* is also required to carry the gradient components from the "radiance space" to the "temperature/humidity space". As T_r is non linear, these gradients are profile dependent, and the computation of T_r uses the latest estimate of the profile as it comes from X at the current iteration of the minimization scheme. This technique is equivalent to incorporating the retrieval scheme into the three-dimensional analysis. The retrieval scheme can, thereby, automatically benefit from all the information available to the analysis: the first-guess and other observations types which are around the satellite observation point. The gradients computed by T_r^* tell the minimization scheme the amount of information to be extracted from each radiance channel for each TOVS observation. This information is profile dependent (or air-mass dependent): to achieve the same result in an Optimal

Interpolation analysis, one would need a set of observation error statistics associated to each observation. The direct use of radiances has the key advantage of avoiding the use of any rather inaccurate information coming from a separate retrieval technique.

The radiative transfer model T_r for TOVS radiances and its adjoint T_r^* are documented in *Eyre* (1991). T_r has been developed not only for three-dimensional and four-dimensional variational assimilations, but also for performing retrievals at ECMWF (1D-VAR). It uses as input the temperature and humidity on 40 "TOVS levels" from 1000 to 0.1 hPa.

The cost function for satellite radiances is defined as:

$$J_{sat} = (R_{mod} - R_o)^t \, O^{-1} \, (R_{mod} - R_o)$$

(R_{mod}: radiance vector computed from X; R_o: observed radiances). The computation of J_{sat} is less straightforward than for most of the other observation types, because the radiance errors may be correlated in the horizontal. Consequently the matrix O is far from being diagonal, and a special numerical technique is needed to compute J_{sat}. A practical solution is based on the following two remarks:

a) J_{sat} can still be split into contributions coming from different satellites (NOAA10, NOAA11, ...) and from different TOVS types (clear, partly cloudy, cloudy).

b)   Assuming the observation error correlation can be split into the product of an inter-channel correlation and a horizontal correlation (usual separability assumption), J_{sat} can be split into contributions coming from the different eigen-vectors of the inter-channel error correlation matrix of TOVS.

Then, instead of inverting O directly, we apply a technique identical to the one described in Appendix A for J_g, except that we have only one parameter (radiance) to cope with. Because of the potential horizontal correlation of TOVS errors, the TOVS operators are organised in a different stream from other observation types. The TOVS sets currently consist of up to 1000 TOVS observations belonging to the same category: same satellite, same cloud-clearing path, same surface type. As with conventional observations, one can multitask the TOVS operators by sending different sets to different processors. The vertical interpolation of thermodynamical variables from the model hybrid levels to the 40 TOVS levels is vectorized on the number of TOVS (up to 1000). The radiative transfer model treats up to 50 profiles at once. To compute the cost function in radiance space, one needs to invert several horizontal correlation matrices, which may be of order 1000 x 1000. Preliminary tests have shown that the radiative computations are quite fast. Several hundred

TOVS operations per second are treated on one processor of the Cray Y-MP8 and the computer time requirements for TOVS operators should not be a problem in the context of three-dimensional or four-dimensional variational analysis.

The NMC variational analysis is quite different in the way TOVS data are used. NMC kept all the observation operators linear, and they use retrieved profiles rather than radiances. The retrieved profiles are moved first to the model levels to be compared with the model. NMC plans to use, in the future, "interactive retrievals" (which are similar to the ECMWF 1D-VAR) in the variational analysis, whereas the ECMWF strategy consists of trying in parallel 1D-VAR in Optimal Interpolation, and radiances in 3D-VAR.

4.   THREE-DIMENSIONAL ANALYSIS EXPERIMENTS

4.1   Experiments with no J_o

The effect of the choice of control variable and balance constraints is most easily seen in the two-dimensional context and in the absence of observations as 1) we then know the solution precisely, and 2) we can then expect to solve the minimization problem under certain conditions in a single step.

We shall take as X_g an initialized analysis from the operational ECMWF assimilation system for 12UTC 14/07/89 and truncated at T21. Since the IFS model used here differs from the operational ECMWF model, this is not a 'balanced' field as far as the IFS is concerned. As starting point for the minimization we shall take the operational uninitialized ECMWF analysis for 12UTC on 15/07/89, again truncated to T21. In the absence of observations and any balance constraints the solution is $X = X_g$, where $J_g = 0$, and $\nabla J_g = 0$. When balance constraints are included the solution is no longer $J_g = 0$, but $\nabla J = 0$ is still the goal (where $J = J_g + J_c$).

The forecast error field has considerable horizontal and vertical variability - as is necessary for proper representation of the forecast error variances. When the model state vector in spectral space X is chosen as control variable, then even for J_g alone, convergence is achieved rather slowly. This is because it is not possible to precondition the minimization sufficiently well simply by specifying the diagonal elements alone (see section 3.3). If, however, x, the departure of the model state variables from the guess, normalized by the forecast error standard deviations, is chosen as control variable then (again, for J_g alone) exact convergence is achieved in a single step of the minimization scheme as in this case the Hessian is diagonal and the correct pre-conditioning may be applied .

16

*For this reason we can treat the full spatial variations of the forecast error variances **only** when x is used as control variable. When **X** is chosen as control variable the forecast error variances are horizontally averaged and vary only in the vertical.* In both cases exact convergence for J_g alone is then achieved in one step of the minimization.

For these two choices of control variable it is useful to examine the effects of including non-linear normal mode initialization (*NNMI*) in the cost function and of adding the J_c term on the balance of the final state. These terms also affect the pre-conditioning as exact convergence can no longer be achieved in a single step of the minimization simply by specifying the diagonal component of the Hessian.

i)   Effect of *NNMI* in the cost function.

As described in section 3.2 (i), for both J_g and J_o it is possible to evaluate the cost function in terms of a distance from the slow manifold -

$$J_g = (NMI(X) - X_g)^t \ P^{-1}(NMI(X) - X_g), \quad J_o = (K(NMI(X)) - d)^t \ O^{-1} \ (K(NMI(X)) - d).$$

This leaves a certain amount of ambiguity in the control variable itself since both X and x in general contain gravity wave components. The minimization will attempt to fit the slow component of X to X_g through modification of the control variable. In practice X seems to reach the slow component of X_g rather quickly, within 5 or 6 iterations. Thereafter the minimization has difficulty in reconstructing the 'fast' components in X_g, and in the case of a 'noisy' X_g may never succeed.

Fig. 1 shows an example where the first guess, X_g, and the starting point for the minimization, X_o, are two initialized analyses 24 hours apart. These have been truncated to T21 from T106. Both fields have been diabatically initialized, whereas the IFS currently only uses adiabatic initialization, therefore the fields are not fully 'balanced' in the IFS sense. The control variable used for this example is X (with horizontally constant forecast error variances). Fig. 1a shows the effect of the NMI on the cost function itself. One iteration of *NNMI* results in slower initial convergence, but with enough passes through the minimization the *NNMI* process can be inverted and X approaches X_g more closely. With two or more iterations of *NNMI* only the 'slow' component of X_g is recovered. Fig. 1b shows the effect on the gradient - the more iterations of *NNMI* the lower the final gradient reached. Finally, Fig. 1c shows the effect on noise control. It is clear that, with only one iteration of *NNMI*, the amount of gravity wave activity increases rapidly. Two iterations

17

of *NNMI* is sufficient to control it to a reasonable level. Comparing with Fig. 1a it is clear that the gravity wave activity only increases when the minimization attempts to fit the fast components of X_g.

As can be seen in Fig. 2, the choice of x as control variable is insensitive to the number of iterations of *NNMI* used. Fig. 2 shows the results obtained by varying the number of *NNMI* iterations between 1 and 8, the lines can almost be overlayed on each other. It is not clear at this stage why it should be so insensitive. Convergence is much improved compared with the use of X as control variable and there is no evidence of increasing noise. As in the X case, convergence is most rapid in the first 5 or 6 iterations of the minimization, suggesting that beyond this the minimization is trying to fit the 'fast' components of X_g.

In summary, the effect of using *NNMI* in the cost function results in X moving towards the slow component of X_g fairly quickly. Use of at least two iterations of *NNMI* ensures that X is kept fairly noise free. Beyond this point (5 or 6 steps of the minimization) little progress is made with X as control variable, although x continues to converge for some time, although at a reduced rate. It should be emphasized that this formulation of the cost function does not ensure a balanced X, merely that the minimization acts in a subspace of X (the slow modes) so that convergence should be quicker.

ii)   Effect of J_c

As mentioned above, use of *NMI(X)* in the cost function will not automatically provide a balanced final state, but it may speed up the initial rate of convergence. Balance considerations are addressed through a further, weak constraint J_c as described in section 3.2 (ii). This J_c is a constraint on the gravity mode tendencies and can be thought of as a progressive *NNMI* applied to X. J_c complements the use of *NMI(X)* in the computation of the J_g and J_o cost functions by introducing a constraint on the fast modes which is absent from these terms.

As can be seen in Fig. 3, initial convergence is generally slower than when *NNMI* is used in the cost function (because minimization is no longer primarily acting on the slow modes) but, when used in combination with *NNMI* in the cost function, convergence is not unduly affected. It does, however, act to progressively dampen the gravity wave tendencies and introduce a degree of balance into the fields. (Note Fig. 3 is not directly comparable to Figs. 1 and 2 as different initial conditions were used for these tests). This result is consistent with the findings of *Courtier and Talagrand* (1990) using a shallow water model.

4.2   Experiments with J_o only

In this set of experiments we try to minimize the observation cost function J_o only, without any guess or constraint cost function. This does not correspond to any realistic scenario as this is equivalent to performing an objective analysis without any guess. The minimization of J_o generally leads to an under-determined global problem with some "local redundancies" in the areas where observations are dense and where the minimization has to achieve "a compromise" in order to fit the model variables to the observed data.

The reason for running this set of "J_o only" experiments is to check the performance of the minimization. The minimization has been preconditioned in the way described in section 3 and used in section 4.1: this preconditioning is based on the J_g cost function only, hoping that it is acceptable for the total cost function. It is then crucial to check to what extent the minimization is efficient in minimizing J_o when no observation operator has been introduced in the preconditioning.

Table 2 summarizes the result of an experiment performed with the T21/L19 version of the IFS, and with the conventional observations for 12Z, 15 July 1989. The starting point of the minimization has been taken equal to the 6h forecast valid for the same date. The control variable x (the normalised model variables, see section 3) has been used for results presented in table 2.

	It 0	It 10	It 25	It 50	Ratio Final/Initial
J_o	88 000	63 000	52 000	43 000	0.49
$\|\nabla J_o\|^2$	1.75×10^6	2.04×10^5	7.70×10^4	3.40×10^4	0.02

Table 2: Observation cost function and square of the norm of its gradient, as a function of the number of iterations in a minimization starting from the 6h forecast valid for the same time as the observations used as the control variable.

The same experiment, performed with the un-normalized control variable X, shows similar results with a somewhat slower convergence. The figures of table 2 indicate that the convergence works reasonably well. The steady decrease of the cost function and its gradient norm also shows that convergence is not reached after 10 iterations, nor even after 25. We do not even know from the results of table 2 to what extent the convergence is reached at iteration 50, as we are no longer in the situation of experiments 4.1 where we knew the value of the cost function at the minimum; this minimum is not 0 because of the local redundancies mentioned before.

One way to know more about the efficiency of the minimization is to vary the initial point of the minimization algorithm: as the final solution is not dependent on the choice of the initial point, we should end up with the same final cost function value if the convergence was perfectly achieved after 50 iterations. Table 3 shows the results corresponding to three different initial points: the 6h forecast as before, the initialized analysis for the same time, and an initialized analysis for 10 days earlier (analysis of 12Z, 5 July 1989).

		It 0	It 10	It 25	It 50	Final/Initial
6h forecast	J_o	88 000	63 000	52 000	43 000	0.49
Analysis 15/07/89	J_o	73 000	56 000	47 000	41 000	0.56
	$\|\nabla J_o\|^2$	7.10×10^5	9.90×10^4	3.80×10^4	1.82×10^5	0.26
Analysis 05/07/89	J_o	600 000	214 000	115 000	65 000	0.11
	$\|\nabla J_o\|^2$	8.13×10^7	2.52×10^6	5.53×10^5	2.26×10^5	0.003

Table 3: Observation cost function and square of the norm of its gradient, as a function of the number of iterations, in a minimization starting from three different initial points.

The interesting feature is the behaviour of the cost function when the minimization is started from a solution which is very far away, i.e. the 10-day old analysis: the initial cost function which is about 8 to 10 times bigger than when one starts from the analysis or the 6h forecast, and it is reduced in 50 iterations to a value which is about the same as in the two other cases. Although the convergence is not perfect, even at iteration 50, this is a good indication of the efficiency of the minimization scheme in an extreme case when one starts very far away from the initial solution and nothing is done to precondition the problem in an optimum way. The square of the gradient norm has been divided by about 300 in this case, which is also a good sign. However, one negative aspect is the increase of the gradient norm from iteration 25 to iteration 50, when the starting point is the analysis of 15/07/89; this may be interpreted as a symptom of poor preconditioning.

All the previous experiments have been run at T21 with fields extracted from the ECMWF operational archives and truncated at T21. The same set of experiments has been run at T42/L19, but instead of truncating the archived fields at T42, a T42 multivariate Optimal Interpolation analysis was run with the ECMWF operational system to provide the T42 fields. The T42 Optimal Interpolation analysis was run with only conventional data, in order to provide a clear reference for variational analysis experiments. Only partial results are available at the time of writing. The T42 results:

- confirm the superiority of x over X for the choice of control variable;

- show a somewhat smaller value for the J_o cost function, compared to T21; this is explained by the higher number of degrees of freedom to fit the observed data;

- show an occasional increase of the gradient norm which can again be interpreted as a sign of bad preconditioning; as in T21 this does not prevent a good reduction of the cost function when the minimization is started from the 10-day persistence analysis.

Experiments with radiance data have been performed separately so far. The radiance operators are the only ones which can lead to a cost function which is far from quadratic. It is interesting to see how the minimization behaves when it is started from a state far from the observed radiances, when there is a maximum risk of finding multiple minima. A T21/L19 experiment has been performed to minimize the radiance cost function corresponding to more than 3000 TOVS observations of 00Z, 9 February 1989. This experiment is even further away from a realistic operational scenario than the experiments with conventional observation, as the minimization is started from the analysis of 12Z, 15 July 1989, which is five months apart! Because we start from a summer situation to assimilate TOVS radiances of a winter situation, the TOVS channel departures are often 10 to 50 K in terms of brightness temperatures, before the minimization. However, the minimization reduces these departures to a few degrees K in less than 30 iterations, in most of the channels. Because of the crudeness of this experiment, this is very encouraging. However, some humidity channels and stratospheric channels show a lack of convergence and require further investigation.

4.3   <u>Full experiments: J_o, J_g and J_c</u>

These experiments correspond to a realistic operational scenario, and can be compared to the operational Optimal Interpolation analysis. Only partial results are available from these experiments at the time of writing, and no complete comparison variational analysis versus Optimal Interpolation has been performed so far.

		It 0	It 10	It 25	It 50	Final/Initial
6h forecast as guess and	J	88 000	66 000	63 000	52 000	0.589
initial point	$\|\nabla J\|^2$	1.75×10^6	1.12×10^5	5×10^3	1.60×10^2	9×10^{-5}

Table 4: Cost function $J = J_o + J_g$, and square of its norm, as a function of the number of iterations in a minimization starting from the guess (i.e. the 6 h forecast).

The results shown in table 4 have been obtained at T21 with the conventional observations of 12Z, 15 July 1989, used together with the corresponding 6h forecast (truncated at T21) as guess. The reduction factor (9×10^{-5}) which has been achieved on $\|\nabla J\|^2$ in fifty iterations is a sign of good convergence. This is confirmed by T42 experiments (not shown). However, no definite conclusion can be drawn before a detailed examination of the resulting variational analyses with strict comparison with Optimal Interpolation analysis. These results were obtained without the use of any balance requirement neither through J_c nor through the use of *NNMI* operator.

On the other hand, the following experiments have been performed with the use of the constraint J_c and the *NNMI* operator, as described in section 3. They are TOVS radiance experiments: the same radiance set is used as before, together with the standard J_g and J_c cost functions, but without any conventional observations. The minimization starts from three different solutions: the 6h forecast (FG), the uninitialized analysis (AN) and the initialized analysis (IA) (all valid for 00Z, 9 February 1989, the same time as the TOVS observations).

	J_o (*10^3) Initial Final	J_g (*10^3) Initial Final	J_c (*10^3) Initial Final	J Final/Initial	$\|\nabla\|^2$ Final/Initial
FG	253 140	6.8 11.9	19.0 4.6	0.56	0.048
AN	244 140	364 72	45.0 8.9	0.34	0.044
IA	244 141	363 73	18.7 10.7	0.36	0.152

Table 5: Initial and final values of the different terms of the cost function after 10 iterations of the minimization. The normalized control variable is used.

The behaviour of all the cost functions is normal in table 5, except the J_g contributions when one starts the minimization from the guess (1st line in table 5): J_g is small before the minimization and stays small after 10 iterations instead of becoming similar to what it is in the two other experiments. This is an indication of

lack of convergence after 10 iterations. Before looking into more details of this problem, it is intended to tune the statistics on observation and forecast errors used in J_o and J_g, as they are very crude in all these experiments.

5. FOUR-DIMENSIONAL ASSIMILATION EXPERIMENTATION

5.1 General description of the experiments

The aim of the 4-D experiments is to evaluate the scientific potential of a four-dimensional variational assimilation technique in which the information contained in the dynamical equations is fully and consistently used together with the observations. They are not intended to represent a realistic operational scenario. A series of four-dimensional experiments have been carried out using the IFS/ARPEGE system at both ECMWF and METEO FRANCE; they are documented in *Thépaut and Courtier* (1991) and *Rabier and Courtier* (1991). They address the following questions:

i) Can the description of the atmosphere at one time be obtained by observations at a different time t_1, or by a chronological series of observations, used consistently with the model dynamics?

ii) Can some variables be described indirectly by the observation of other variables, or by a time series of other variables, through the model dynamics?

iii) Can some spatial scales, or typical features, be described indirectly by the observation of other scales or features, or a time series of them, through the model dynamics?

iv) Can we get a model trajectory reasonably free of gravity waves and still accurately fit the observations?

Three four-dimensional experiments by *Thépaut and Courtier* (1991) and *Rabier and Courtier* (1991) are summarized in section 5.2. The experiments have been performed with a T21, 19 level model, with no stretching and no pole rotation, without any physical parametrizations and without specific humidity. They are all performed with simulated observations in the form of spectral coefficients. The observations are produced by a run of the IFS/ARPEGE forecast model which is also used as the assimilating model so the experiments are identical twin experiments. In all the experiments the model generated "observations" $d(t)$ are assumed to be perfect, as they are part of the preliminary reference run $X_{ref}(t)$. Starting from a different model trajectory $X(t)$ reasonably far away from $X_{ref}(t)$, we try to minimize the distance:

$$J(X) = \sum_{i=o}^{n} < K(X(t_i)) - d(t_i), K(X(t_i)) - d(t_i) >.$$

However, since the simulated observations $d(t)$ are model state variables, $d(t_i) = X_{ref}(t_i)$, the operator K is the identity operator so $J(X)$ is simply:

$$(1) \qquad J(X) = \sum_{i=o}^{n} < X(t_i) - X_{ref}(t_i), X(t_i) - X_{ref}(t_i) >$$

where n is the number of time steps, $<, >$ is the scalar product and $X(=X(t_o))$ is the control variable. The scalar product is defined by:

(2)

$$<X, X> = 1/2 \int_0^1 \int \int_\Sigma \left(\nabla \Delta^{-1} \xi \cdot \nabla \Delta^{-1} \xi + \nabla \Delta^{-1} \delta \cdot \nabla \Delta^{-1} \delta + R_a T_r (\log p_s)^2 + \frac{C_p}{T_r} T^2 \right) d\Sigma \frac{dp}{d\eta} d\eta$$

which is a quadratic invariant of the linearised primitive equations in the vicinity of a state of rest defined by a constant reference temperature T_r; ξ and δ are the vorticity and divergence fields. This scalar product is also used for defining the metric of the minimization space.

The minimization algorithm used in these experiments is a mixed quasi-Newton conjugate gradient type (*Buckley and Lenir*, 1983). More information is given in section 6 on minimization algorithms.

5.2   Results of three typical experiments

5.2.1   *Inference of the initial state from the final state of a 6 hour forecast*

One assumes that the observations consist of the complete atmospheric fields (vorticity, divergence, temperature and surface pressure) at time t_n which is the end of a 6 h assimilation period:

$$t_n = t_o + 6h$$

Then, one tries to reconstruct the model state at time t_o using these observations at time t_n, i.e. one tries to "invert" the 6 h forecast. The cost function to be minimized is:

$$J(X) = <X(t_n) - X_{ref}(t_n), X(t_n) - X_{ref}(t_n)>$$

The control variable is $X = X(t_o)$, and we start the minimization from an initial state $X_{in}(t_o)$.

The observations and the initial state of the minimization have been constructed in the following way: the IFS model has been integrated for 48 h from the ECMWF operational initialised analysis of 12 July 1989, 12Z. The resulting forecast valid for 14 July 1989, 12UTC is taken as the initial state $X_{ref}(t_o)$ of the reference run. The initial point of the minimization $X_{in}(t_o)$ has been produced in the same way by integrating for 48 h the ECMWF operational 24 h forecast valid for 12 July 1989, 12UTC. The four-dimensional assimilation experiment has been performed on the 6 h time period [14 July 1989 12UTC, 14 July 1989 18UTC]. Such a generation process of the fields allows one to get rid of most of the transients due to the inconsistency between the adiabatic IFS model and the operational analysis and to start the minimization process with a difference between the initial point $X_{in}(t_o)$ and the final solution $X_{ref}(t_o)$ corresponding to the order of magnitude of a 24 h forecast error.

Fig. 4 shows the variation of the cost function $J(X)$ with the number of iterations in the minimization process. Since we know both the solution $(X = X_{ref}(t_o))$ and the value of the minimum $(J(X) = 0)$, in this idealized experiment, it is easy to check that the experiment works. After 30 iterations, the cost function is reduced by 6 orders of magnitude and the norm of the gradient by 3 orders of magnitude. The distance between X and $X_{ref}(t_o)$ at the initial time of the assimilation period is interesting to examine since it measures the ability of the minimization to recover the reference field through the model dynamics. This distance is presented in Fig. 5: it decreases by about 3 orders of magnitude in 30 iterations. This decrease is meteorologically acceptable as shown in Fig. 6, which presents the difference between the reference and the analysed vorticity (left) and temperature (right) for the Northern Hemisphere at 500 hPa. Comparisons between top and bottom panels show to what extent the difference $X(t_o) - d(t_o)$ has been reduced by the minimization for vorticity and temperature.

This experiment validates the numerical feasibility of a 6 h variational assimilation in the extreme (most difficult) context where information is available only 6 hours later than t_o. This particular experiment has been performed without any horizontal diffusion. Other experiments made with varying diffusion coefficients showed that an increased diffusion coefficient makes the inversion of the model more difficult.

5.2.2 *Reconstruction of wind field by assimilating mass observations*

Several four-dimensional experiments have been performed in which only some of the meteorological fields were 'observed'. In one of them, the full mass field is assumed to be observed every hour on a 6 hour period (temperature + surface pressure). Fig. 7 shows to what extent the vorticity field is reconstructed at the beginning of the assimilation period. The vorticity difference fields at 500 hPa between the model solution

and the reference state are dramatically reduced in the mid-latitudes, but the minimization is inefficient in the tropical belt.

The result shown in Fig. 7 has been performed with non-linear normal mode initialization at time t_o, but the same experiment performed without normal mode initialization gives similar results. They both show the ability of the four-dimensional assimilation to reconstruct the vorticity field from the mass field in mid-latitudes but not in the tropics.

As far as the reconstruction of divergence is concerned, an important reduction of the differences has also been noticed. This reduction is likely to be due to the sequential observation of the surface pressure, which gives indirectly through its tendency some information on the divergence.

5.2.3 *Reconstruction of a baroclinic wave from the evolution of the zonal flow*

The reference run $X_{ref}(t)$ in this experiment is an academic situation described in *Simmons and Hoskins* (1978), where a wave pattern develops and interacts with the basic zonal flow. The purpose of this particular experiment is to assess the potential of a four-dimensional assimilation scheme in the context of a rapidly developing weather system. In practice, the reference run $X_{ref}(t)$ is constructed as follows:

i) A baroclinic wave on the sphere is set up as in *Simmons and Hoskins* (1978): the basic flow is chosen symmetric around the equator with a baroclinic zone centered at 45° latitude and characterized by a horizontal temperature gradient. This flow is then perturbed by the most unstable mode at zonal wave number six. This mode is scaled so as to give 1 hPa maximum amplitude for the surface pressure.

ii) The IFS model is integrated for 15 days starting from the previous situation. In the integration the perturbation develops like in a real weather system. After nine days of growth, the minimum value of the surface pressure has dropped from 1008 hPa to 972 hPa; one can then see a decay of the wave. The most intense cyclogenesis occurs between days 6 and 8 and the assimilation experiment has been made on the 24 h period starting at day 6.

In this assimilation experiment, the observations are all zonal wave numbers except $m = 6$, which characterizes the synoptic weather system that we try to reconstruct through the assimilation. As the initial point of the minimization the reference state at day 4 has been used. In summary we try to recover the wave number 6 of the reference state at day 6, from the observations of other wave numbers in the period [day 6, day 7] starting from the reference state at day 4.

Fig. 8 clearly shows the decrease of the cost function, and the paradoxical increase in the distance between the result of the minimization and the reference. The explanation can be seen in Fig. 9. The eddy intensity of the result of the assimilation process is approximately correct, but the pattern is totally out of phase. The dynamics included in the assimilation method are able to infer the development of the system from the temporal evolution of the zonal part of the flow. However, due to the symmetry of the problem, the feedback on the zonal part is independent of its location. The location of the system given by the initial point of the minimization (day 4) is kept unchanged (i.e. wrong by two days).

This result shows an example of how information present in the model dynamics (development of the system) is well used by the four-dimensional assimilation, and an example of information which is lacking in the observations and which cannot be inferred. In another experiment, where part of the surface pressure is included in the observations (including for $m = 6$) the assimilation manages to reconstruct the system at the right place.

5.3   Summary of the results in mid 1991 and short term plans

Several four-dimensional experiments performed with simulated observations and an adiabatic version of the IFS at T21/L19, have demonstrated the potential of a four-dimensional assimilation scheme for using consistently the model dynamical information together with the observations. The inversion of the model (5.2.1) on a 6 h period worked very well. More iterations of the minimization schemes were needed when the period was 12 or 24 h (results not shown). This is an indication that the preconditioning of the minimization will be the key efficiency issue when implementing a four-dimensional operational scheme. The control of gravity waves works if we use a technique similar to *Courtier and Talagrand* (1990) and described in this paper (see section 3.2 and 4.1). The ability of the four-dimensional assimilation to recover some fields or some features through the model dynamics has been demonstrated in examples such as the two experiments presented in 5.2.2 and 5.2.3.

The next step, starting in mid 1991, is to perform four-dimensional variational assimilation with real observations, using the observation operators documented in section 3. Then, the effects of including physical parametrizations in the model will have to be considered. Once these two aspects have been studied, some attention can be given to the practical aspects related to a possible operational scenario: particularly computer cost in relation to the preconditioning problem mentioned before.

6.   MINIMIZATION ALGORITHMS

No developments in minimization algorithms have been made at either ECMWF or METEO FRANCE. The results presented before have been obtained by using minimization packages from Institut National de la

Recherche en Informatique et Automatique (INRIA) and used as "black boxes". These packages, called "M1GC2" and "M1GC3", are different versions of a mixed quasi-Newton/conjugate gradient algorithm (*Buckley and Lenir*, 1982). *Navon and Legler* (1987) provide a detailed review of the minimization packages available for the meteorological problems and a description of the Buckley and Lenir algorithm.

The first descent steps are performed following the quasi-Newton method, as long as the memory allocated for the storage of the approximation of the inverse of the Hessian is not full. As soon as there is no further memory, it switches to a conjugate-gradient method preconditioned by the latest approximation of the Hessian inverse.

In addition to the two packages from INRIA, M1GC2 and M1GC3, another package is available at ECMWF and METEO FRANCE: LBFGS (Limited Broyden Fletcher Goldfarb Shanno) coded by L Liu and J Nocedal at Northwestern University. Some experiments have been carried out at ECMWF in order to compare the different packages regarding computer time, memory size and accuracy of the solution. We have chosen a minimization problem whose solution is perfectly known: the four-dimensional assimilation experiment consisting of inverting the model at T21/19L and documented in 5.2.1.

In the LBFGS method, quasi-Newton corrections are stored separately and when the available memory is used up, the quasi-Newton matrix is updated by dropping the oldest information and replacing it with the newest information available to produce a step direction. Therefore the quasi-Newton approximation of the inverse Hessian matrix is continuously updated. Three different options of LBFGS have been considered, but only the one in which the Hessian matrix is initialized by the identity matrix and is rescaled automatically at each iteration actually worked.

The results showed first that M1GC2 and M1GC3 have completely identical performance. Then LBFGS and M1GC3 have been compared under two sets of conditions: 11 updates or 38 updates of the initial Hessian matrix. The different results are shown in Fig. 10 in terms of the decrease of cost function with respect to the number of iterations. The four algorithms perform in a very similar way up to iteration 20 and even up to iteration 50 if we exclude M1GC3 with only 11 updates which is slightly less efficient above iteration 20. The algorithms with 38 updates use up to 3 million words of memory, three times more than the algorithms with 11 updates. The CPU time of the four algorithms is the same, except the LBFGS algorithm with 38 updates is slightly faster than the others.

7.   CONCLUSION

Based on the concept of an adjoint operator, a variational assimilation has been developed in the context of the IFS/ARPEGE system which is common to ECMWF and METEO-FRANCE. This variational assimilation

can work in three-dimensional or four-dimensional mode; it can use real observations as well as simulated ones in the form of spectral or grid-point fields. The code has been developed and tested for all the conventional observations currently used in the ECMWF operational analysis, and also for the TOVS clear radiances. Two options have been developed for choosing the control variable: in the first one the control variable is the ensemble of model variables; in the second one it consists of the model variables normalized by the forecast error standard deviations. The mass/wind balance is treated by two developments: one is the use of normal mode initialization and its adjoint, the other one is a penalty term which keeps the tendency of gravity waves small.

Several three-dimensional variational experiments have been performed with a preconditioning of the minimization which has the property to be optimal for J_g. These experiments performed with J_o only, and/or with $J_o + J_g$, with or without balance constraints, have demonstrated that the minimization was working reasonably well in all the configurations. This is an indication that the real size variational analysis (which obviously contains J_o, J_g and J_c) is feasible in practice, at a reasonable computer cost, without much extra work on optimizing the minimization.

This last point still needs to be fully proved before planning the details of an operational three-dimensional variational analysis. It is hoped that this step can be reached in Autumn 1991. The work to be done then in order to reach an operational implementation (one year later) consists of:

-   technical developments in order to cope with operational resolutions; this implies Input/Output developments and technical choices, especially for the minimization schemes;

-   tuning the different elements of the variational analysis, especially all the statistics;

-   decide exactly what data to use/not to use (e.g. implement the boundary layer operators which are not developed yet; - decide what TOVS channels to use);

-   perform the final experimentation/evaluation of the three-dimensional variational analysis.

The operational implementation of a three-dimensional variational analysis is certainly not the last step of the project. Further developments are likely to occur in this operational scheme, such as using more observations, or improving the use of current observations. Also more detailed studies on the operational feasibility of a four-dimensional variational scheme will be performed.

A series of four-dimensional assimilation experiments have been performed with simulated data in spectral space. They showed that it is possible to "invert" the forecast model through a variational four-dimensional assimilation performed in a 6 hour period (t_o, t_1): in other words it is possible to reconstruct the model at time t_o from observing the model state at time t_1. This inversion is becoming more and more difficult when diffusion is increased in the model, and/or when the time period (t_o, t_1) is increased from 6 to 12 or 24h. In other experiments it has been shown that the information available in the model dynamics is sufficient to reconstruct several model fields from the observation of other model fields. The reconstruction of an idealized baroclinic wave has also been simulated with promising results.

In general, $J_g = (\underline{X} - \underline{X}_g)^t \, \underline{P}^{-1} \, (\underline{X} - \underline{X}_g)$,

where $\underline{X}$ and $\underline{X}_g$ are vectors of the spectral coefficients of the model state variables.

Separability will be assumed in the horizontal and vertical, so that

$$J_g = (X-X_g)^t \, \underline{V}^{-1} \, \underline{H}^{-1} \, (X-X_g)$$

where $\underline{V}$ is a correlation matrix and $\underline{H}$ is a covariance matrix. Normalizing $(X - X_g)$ by the forecast error standard deviations one obtains,

$$J_g = \left(\frac{(X - X_g)}{\sigma}\right)^t \, \underline{V}^{-1} \underline{h}^{-1} \left(\frac{(X - X_g)}{\sigma}\right)$$

where $\underline{V}^{-1}$, and $\underline{h}^{-1}$ are both now correlation matrices, containing within them the vertical and horizontal

components respectively. Note that, in general, $\underline{H}^{-1} \neq \left(\frac{1}{\sigma}\right)^t \underline{h}^{-1} \left(\frac{1}{\sigma}\right)$, the relation between $\underline{h}$ and $\underline{H}$ is not

simple. From now on we shall work entirely in terms of $\underline{h}$.

Let, $x = \left(\frac{X - X_g}{\sigma}\right)$ $\qquad$ then, $J_g = x^t \, \underline{V}^{-1} \, \underline{h}^{-1} \, x$ $\qquad$ and $\nabla_x J_g = 2 \, \underline{V}^{-1} \, \underline{h}^{-1} \, x$.

Let, $\quad \underline{F} = \underline{h}^{-1} \, x, \quad \underline{G} = \underline{V}^{-1} \, \underline{F}, \quad$ then $J_g = x^t \, \underline{G}, \quad \nabla_x J_g = 2\underline{G}$

The vector x contains the normalized departures of the model state variables from the guess field

$$x = \begin{pmatrix} \xi \\ \delta \\ T \\ q \\ \ln p_s \end{pmatrix}, \text{ where for example } \xi = \begin{pmatrix} \xi_1^{level1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \xi_N^{level1} \\ \xi_1^{level2} \\ \cdot \\ \cdot \\ \cdot \\ \xi_N^{level2} \\ \cdot \\ \cdot \\ \xi_1^{levelL} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \xi_N^{levelL} \end{pmatrix}$$

31

$\underline{\xi}$, $\underline{\delta}$, $\underline{T}$, $\underline{q}$ contain the (N) spectral coefficients of the normalized departures of the model state variables from the guess field at each model level.

If we assume that the cross correlations between errors in model variables are zero, we may define a block diagonal matrix $\underline{\underline{h}}^{-1}$

$$\underline{\underline{h}}^{-1} = \begin{pmatrix} \underline{\underline{h}}_{\xi} & 0 & 0 & 0 & 0 \\ 0 & \underline{\underline{h}}_{\delta} & 0 & 0 & 0 \\ 0 & 0 & \underline{\underline{h}}_{T} & 0 & 0 \\ 0 & 0 & 0 & \underline{\underline{h}}_{q} & 0 \\ 0 & 0 & 0 & 0 & \underline{\underline{h}}_{\ln p_s} \end{pmatrix}$$

where $\underline{\underline{h}}_{\xi}$ is an NLEV x NLEV matrix of block diagonal components $\underline{\underline{h}}_{\xi l}$, where l refers to the model level

$$\underline{\underline{h}}_{\xi} = \begin{pmatrix} \underline{\underline{h}}_{\xi 1} & 0 & 0 & \cdot & \cdot \\ 0 & \underline{\underline{h}}_{\xi 2} & 0 & \cdot & \cdot \\ 0 & 0 & \underline{\underline{h}}_{\xi 3} & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \end{pmatrix}$$

With similar matrices for $\underline{\underline{h}}_{\delta}$, $\underline{\underline{h}}_{T}$, $\underline{\underline{h}}_{q}$. For $\underline{\underline{h}}_{\ln p}$, there is only a single level.

Since $\underline{\underline{h}}^{-1}$ is operating on $\underline{x}$, which has been normalised by the forecast errors, σ, it is reasonable to assume that $\underline{\underline{h}}^{-1}$ is homogeneous ie $\underline{\underline{h}}_{\xi 1} = \underline{\underline{h}}_{\xi 2} = \ldots = \underline{\underline{h}}_{\xi l}$ which are diagonal in spectral space. We shall further assume that

$$\underline{\underline{h}}_{\xi l} = \underline{\underline{h}}_{\delta l} = \underline{\underline{h}}_{Tl} = \underline{\underline{h}}_{ql} = \underline{\underline{h}}_{\ln ps} = \underline{\underline{h}}_{n}$$

where for each level $\quad \underline{\underline{h}}_{n} = \begin{pmatrix} h_1 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & h_2 & 0 & & & \\ 0 & 0 & h_3 & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ \cdot & & & & & h_N \end{pmatrix}$

a matrix of dimension N: the number of spectral coefficients.

$$\underline{F} = \underline{h}^{-1} \, x = \begin{pmatrix} \underline{h}_\xi & 0 & 0 & 0 & 0 \\ 0 & \underline{h}_\delta & 0 & 0 & 0 \\ 0 & 0 & \underline{h}_T & 0 & 0 \\ 0 & 0 & 0 & \underline{h}_q & 0 \\ 0 & 0 & 0 & 0 & \underline{h}_{\ln p_s} \end{pmatrix} \begin{pmatrix} \xi \\ \delta \\ T \\ q \\ lnps \end{pmatrix}$$

$$= \begin{pmatrix} \underline{h}_\xi \, \xi \\ \underline{h}_\xi \, \delta \\ \underline{h}_T \, T \\ \underline{h}_q \, q \\ \underline{h}_{lnps} \, lnps \end{pmatrix} = \begin{pmatrix} \xi^h \\ \delta^h \\ T^h \\ q^h \\ lnps^h \end{pmatrix}$$

where
$$\xi^h = \begin{pmatrix} \begin{pmatrix} h_1 \ \xi_1 \\ h_2 \ \xi_2 \\ \cdot \\ \cdot \\ h_N \ \xi_N \end{pmatrix}_{level \ 1} \\ \cdot \\ \cdot \\ \cdot \\ \begin{pmatrix} h_1 \ \xi_1 \\ h_2 \ \xi_2 \\ \cdot \\ \cdot \\ h_N \ \xi_N \end{pmatrix}_{level \ L} \end{pmatrix}$$

with similar expressions for δ^h, T^h, q^h and $\ln p_s^h$.

Let
$$\underline{V}^{ij} = \begin{pmatrix} V_{11}^{ij}I & V_{12}^{ij}I & \ldots & V_{1L}^{ij}I \\ V_{21}^{ij}I & V_{22}^{ij}I & \ldots & \\ \cdot & & & \\ \cdot & & & \\ V_{L1}^{ij}I & \ldots & & V_{LL}^{ij}I \end{pmatrix}.$$

Where I is a unit matrix of dimension N and $i,j = \xi,\delta,T,q,\ln p_s$

$$\begin{pmatrix} \underline{V}^{\xi\xi} & \underline{V}^{\xi\delta} & \underline{V}^{\xi T} & \underline{V}^{\xi q} & \underline{V}^{\xi \ln p_s} \\ \underline{V}^{\delta\xi} & \underline{V}^{\delta\delta} & \underline{V}^{\delta T} & \underline{V}^{\delta q} & \underline{V}^{\delta \ln p_s} \\ \underline{V}^{T\xi} & \underline{V}^{T\delta} & \underline{V}^{TT} & \underline{V}^{Tq} & \underline{V}^{T \ln p_s} \\ \underline{V}^{q\xi} & \underline{V}^{q\delta} & \underline{V}^{qT} & \underline{V}^{qq} & \underline{V}^{q\ln p_s} \\ \underline{V}^{\ln p_s, \xi} & \underline{V}^{\ln p_s,\delta} & \underline{V}^{\ln p_s, T} & \underline{V}^{\ln p_s, q} & 1 \end{pmatrix} .$$

For simplicity it is assumed that the cross correlations are again zero - the off diagonal blocks $\underline{V}^{ij} = 0$ for $i \neq j$.

Then $\quad \underline{G} = \underline{V}^{-1} \, \underline{F}$

$$\underline{G} = \begin{pmatrix} \underline{V}^{\xi\xi} \, \underline{\xi}^h \\ \underline{V}^{\delta\delta} \, \underline{\delta}^h \\ \underline{V}^{TT} \, \underline{T}^h \\ \underline{V}^{qq} \, \underline{q}^h \\ \ln p_s^{\,h} \end{pmatrix}$$

$\nabla_x J_g = 2\underline{G}$ and $J_g = x^t \, \underline{G}$.

All that remains is to define the matrices $\underline{V}^{\xi\xi}$, $\underline{V}^{\delta\delta}$, $\underline{V}^{TT}$, $\underline{V}^{qq}$

and $\quad \underline{h}_n = \begin{pmatrix} h_1 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & h_2 & 0 & & & \\ 0 & 0 & h_3 & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ \cdot & & & & & h_N \end{pmatrix} .$

Specification of the 'horizontal' error correlation matrix $\underline{h}$

The assumption built into the horizontal prediction error correlation matrix $\underline{h}_n$ is that the structure functions

(on the gaussian grid) are homogeneous and isotropic on the sphere. In physical space, we have then taken: $h(r) = e^{-\left(\frac{r^2}{2a^2}\right)}$

where r is the distance between two points on the sphere and a is the radius of influence (600 km). This hypothesis of homogeneity leads to a diagonal error correlation matrix in the spectral space (*Courtier*, 1987).

Specification of the vertical error correlation matrix $\underline{V}$

$\underline{V}$ is a block diagonal matrix with block elements square matrices of dimension NLEV, one block for each of the 3-D state variables. As a starting point these vertical correlation matrices have been closely modelled on those used operationally within the OI analysis. The vorticity correlations have the same form as those used for streamfunction in the OI, the divergence follow those of velocity potential in the OI. Temperature correlations have been derived from the height correlations used in the OI. Specific humidity follows that of the OI form for relative humidity.

These represent a crude first step and will be refined in the future. In particular, deriving temperature correlations from those of height introduces undesirable 'negative lobes' in the correlation structure. Specific humidity correlations need to be properly derived from those of relative humidity, or determined in their own right.

APPENDIX B

B.1 *Calculation of forecast errors within the OI*

Forecast errors are estimated within the ECMWF operational OI assimilation as follows: -

Estimates of the rms forecast errors of u, v and h (E^{fc}_u, E^{fc}_v, E^{fc}_h) have been statistically determined through verification of the assimilation scheme in data dense regions. (*Hollingsworth and Lönnberg*, 1986). Smooth, **global**, three-dimensional fields have been generated from these results. The independent $E^{fc}_{u,v}$ are used close to the equator. Away from the equator, they are calculated as:

$$E^{fc}_{u,v} = \frac{g}{\sqrt{2Lf}} E^{fc}_{u,v}$$

where L is a horizontal length scale and is the Coriolis parameter.

Modification of the mean errors $E^{fc}_{u,v,h}$ is necessary to reflect spatial and temporal variation in data density. This is carried out through the following steps: -

i) A global three-dimensional field of analysis errors $E^{a}_{u,v,h}$ is available as a by product of the OI analysis. These are used to modify the climatological forecast errors to reflect data availability for the specific forecast.

$$e^{a} = \frac{1}{3N} \sum_{i=1}^{3} \sum_{j=1}^{N} \frac{E^{a}_{ij}}{E^{fc}_{ij}}$$

where $i = 1,2,3$ is u,v,h respectively, and j is a sum over levels. e^{a} is a global 2-D field.

ii) Climatological variances are available from station statistics e.g. Oort and Rasmussen. These are used to provide $E^{c}_{u,v,h}$, which provides a climatological error factor.

$$e^{c} = \frac{1}{3N} \sum_{i=1}^{3} \sum_{j=1}^{N} \frac{E^{c}_{ij}}{E^{fc}_{ij}}$$

iii) Error growth factor. Let e^{a} grow to the error level of a random state in time Δt_{rand}.

$$e^{mod} = e^{a} + \frac{\Delta t_{fc}}{\Delta t_{rand}} \left(\sqrt{2}\, e^{c} - e^{a} \right)$$

37

Δt_{rand} = 6 days in mid-latitudes, 2 days in tropics. Δt_{fc} = 6 hours.

iv)   Apply a 2-D filter to e^{mod}.

v) $E^{fc} = e^{mod} \times E^{fc}$.

Finally, these forecast errors are saved during the ECMWF OI analysis for use within the variational analysis.

B.2   *Treatment of the forecast errors within the IFS*

As described in A.1 an estimate of the rms forecast errors of $E^{u,v,h}$ is available from the operational OI analysis on a 6° x 6° lat/long grid at 7 pressure levels (1000, 500, 300, 200, 100, 50 and 10 hPa).

i)   Horizontally interpolate $E_{u,v,h}$ to the gaussian grid of the IFS.

ii)   Vertically interpolate the $E_{u,v,h}$ to the η-levels of the IFS, using the first-guess surface pressure to define the levels.

iii)   Generate error fields for the variables used in the IFS: contra- and co-variant components of the wind, T, $\ln p_s$, and q. rms forecast error fields used in the IFS

Contra and co-variant wind.

$$\sigma_{u*,v*} = E_{u,v} cosine(lat)$$

Temperature.

$$\sigma_T = \alpha_\eta \|E_h\|$$

The vertical covariance matrix, described in Appendix A, is used to provide the vertical structure of the temperature variances (α_η). Horizontal structure is introduced by using these to scale the normalized height errors $\|E_h\|$ (i.e. the height error field from the OI is used to provide the horizontal structure for the temperature errors).

Surface pressure.

$$\sigma_{p_s} = \rho g E_{h_s}$$

$$\sigma_{\ln(p_s)} = \frac{\rho g E_{h_s}}{p_s}$$

Specific humidity.

38

$$\sigma_q = \frac{(1 - q)^2 e_s R_d}{(p - e_s)R_v}\left(\frac{\Delta r}{100}\right)$$

where currently $\Delta r = 10\%$ (which may be optimistic).

For $p < 200$ hPa, $E_q = 1.25 \times 10^{-6}$ kg/kg.

Options exist for further processing of the rms errors, such as setting σ's to constant values over η-levels, and spectral smoothing of $(1/\sigma)$. The latter is necessary to avoid too much aliasing in the calculation of the cost function.

<u>References</u>

Andersson, E., A. Hollingsworth, G. Kelly, P. Lönnberg, J. Pailleux and Z. Zhang, 1991: Global observing system experiments on operational statistical retrievals of satellite soundings. Mon.Wea.Rev., <u>119</u> No. 8, 1851-1864.

Buckley, A. and A. Lenir, 1983: QN-like variable storage conjugate gradients. Mathematical Programming, <u>27</u>, 155-175.

Courtier, P., 1987: Application du contrôle optimal à la prévision numérique en météorologie. Thèse de doctorat d'état de l'université Paris 6. Available from METEO FRANCE/Toulouse.

Courtier, P. and J.F. Geleyn, 1988: A global spectral model with variable resolution - application to the shallow water equations. Q.J.R.Met.Soc., <u>114</u>, 1321, 1326.

Courtier, P. and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation - Part II. Numerical results. Q.J.R.Met.Soc., <u>113</u>, 1329-1347.

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

Durand, Y., 1985: The use of satellite data in the French high resolution analysis. Proceedings of the ECMWF Workshop on "High resolution analysis", ECMWF, Reading, U.K., May 1985, 89-127.

Eyre, J.R., 1991: A fast radiative transfer model for satellite sounding systems. ECMWF Technical Memorandum No. 176.

Gallimore, R.G. and D.R. Johnson, 1986: A case study of GWE satellite data impact on GLA assimilation analysis of two ocean cyclones. Mon.Wea.Rev., <u>114</u>, 2016-2032.

Gilchrist, A., 1982: JSC Study Conference on Observing system experiments. Exeter, 19-22 April 1982. GARP report No. 4 on Numerical Experimentation Programme.

Halem, M., E. Kalnay, W.E. Baker and R. Atlas, 1982: An assessment of the FGGE satellite observing system during SOP-1. Bull.Amer.Met.Soc., <u>63</u>, 407-426.

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, <u>38A</u>, 111-136.

Kelly, G., E. Andersson, A. Hollingsworth, P. Lönnberg, J. Pailleux and Z. Zhang, 1991: Quality control of operational physical retrievals of satellite soundings data. Mon.Wea.Rev., <u>119</u>, No. 8, 1866-1880.

Kelly, G., 1990: Quality control of satellite data in the ECMWF data assimilation system. BMRC Research Report No. 27 containing the proceedings on the Australian Workshop on Data assimilation systems, September 1990. Available from BMRC, Melbourne, Australia.

Le Dimet, F.X. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus, <u>38A</u>, 97-110.

Lewis, J.M. and J.C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. Tellus, <u>37A</u>, 309-322.

Lorenc, A.C., 1981: A global three-dimensional multivariate statistical analysis scheme. Mon.Wea.Rev., <u>109</u>, 701-721.

Lorenc, A.C., 1988: Optimal non-linear objective analysis. Q.J.R.Met.Soc., 114, 205-240.

Navon, I.M. and D.M. Legler, 1987: Conjugate-gradient methods for large-scale minimization in minimization in meteorology. Mon.Wea.Rev., 115, 1479-1502.

Pailleux, J., 1988: Variational analysis: Use of observations - example of clear radiances. Proceedings of the ECMWF Seminar on Data assimilation and the use of satellite data, Reading, 5-9 September, 1988 (available from ECMWF).

Pailleux, J., 1989: Design of a variational analysis. Organisation and main scientific points. ECMWF Technical Memorandum No. 150.

Pailleux, J., 1990. The development of a global variational assimilation scheme. BMRC Research Report No. 27 containing the proceedings of the Australian Workshop on Data assimilation systems, September 1990. Available from BMRC, Melbourne, Australia.

Parrish, D.F. and J.C. Derber, 1991: The NMC spectral statistical interpolation analysis system. To appear in Mon.Wea.Rev. Also available from NMC/Washington as Technical Note.

Rabier, F. and P. Courtier, 1991: Optimal control applied to the study of baroclinic instability. To be submitted to Q.J.R.Met.Soc. Also available from METEO-FRANCE/EERM/Toulouse as Technical Note.

Simmons, A.J. and B.J. Hoskins, 1978: The life cycles of some non-linear baroclinic waves. Q.J.R.Met.Soc., 35, 414-431.

Simmons, A.J. and D. Burridge, 1981: An energy and angular momentum conserving vertical finite difference scheme and hybrid vertical coordinate. Mon.Wea.Rev., 109, 758-766.

Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: Theory. Q.J.R.Met.Soc., 113, 1313-1328.

Talagrand, O., 1988: Four-dimensional variational assimilation. Proceedings of the ECMWF Seminar on Data assimilation and the use of satellite data, Reading, 5-9 September, 1988 (available from ECMWF).

Thépaut, J.N. and P. Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive equation model. To appear in Q.J.R.Met.Soc. Also available from ECMWF as Technical Memorandum 178.
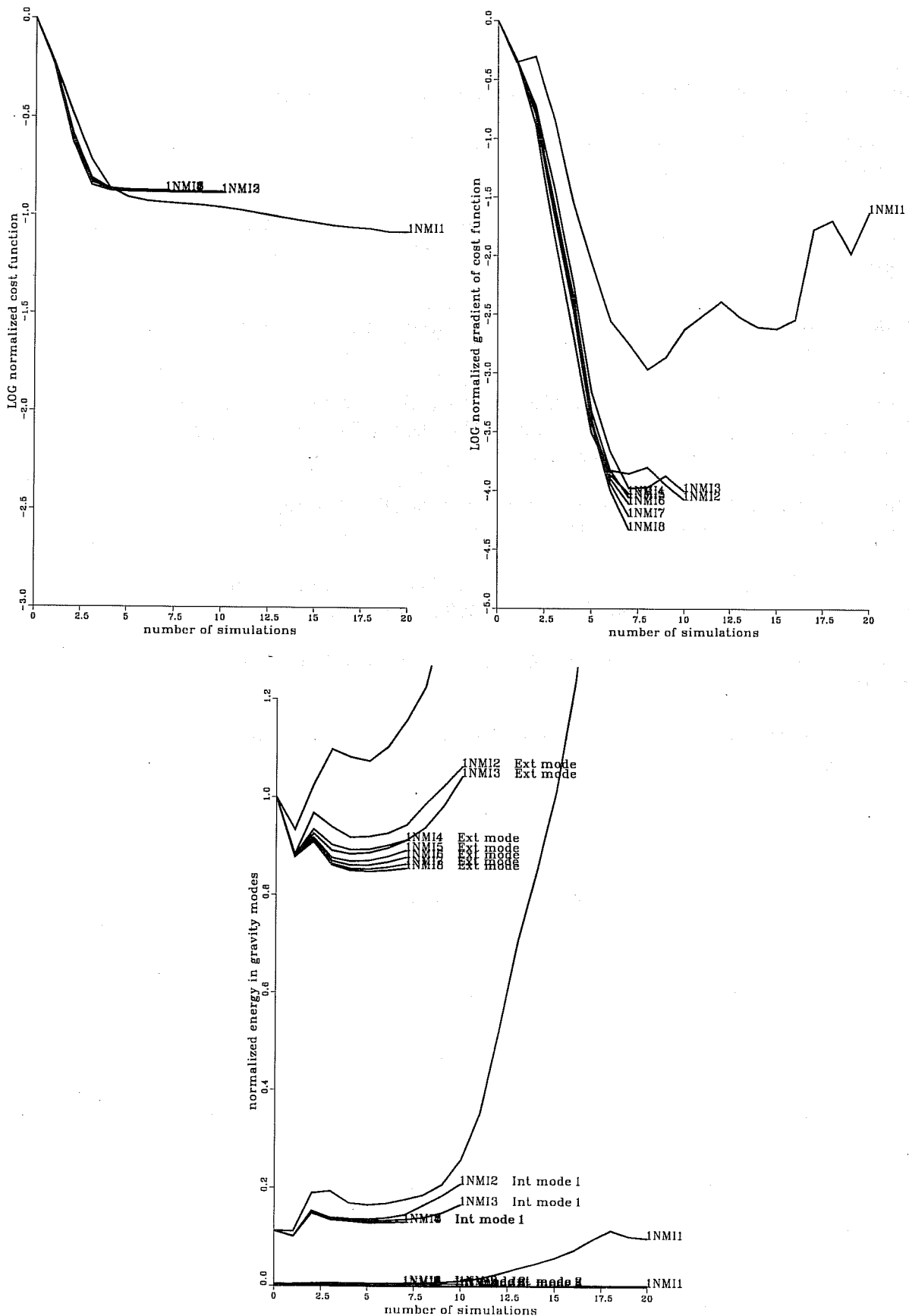
Fig. 1   Convergence for the guess cost function J_g alone with the model variable vector X as control variable

    a)   decimal logarithm of the ratio of J_g at each simulation (of the minimization scheme) to its initial value. The number following "NMI" in the line labels indicates the number of iterations of $NNMI$ used in the computation of J_g.

    b)   As a) but for the gradient of J_g.

    c)   Decimal logarithm of the energy in gravity modes normalized by its initial value, at each simulation.
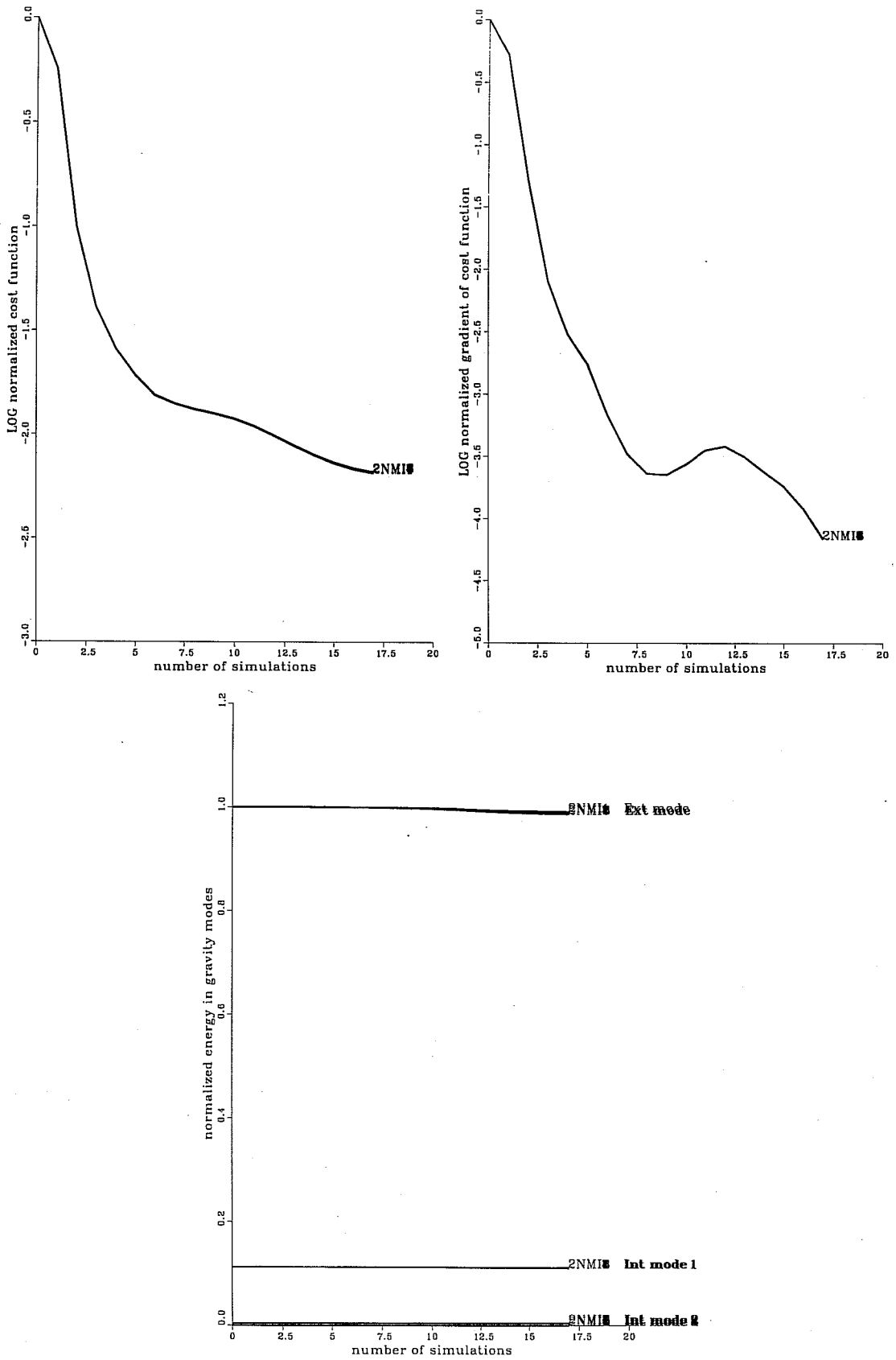
Fig. 2   As Fig. 7 but for the normalized departures to guess x as control variable.
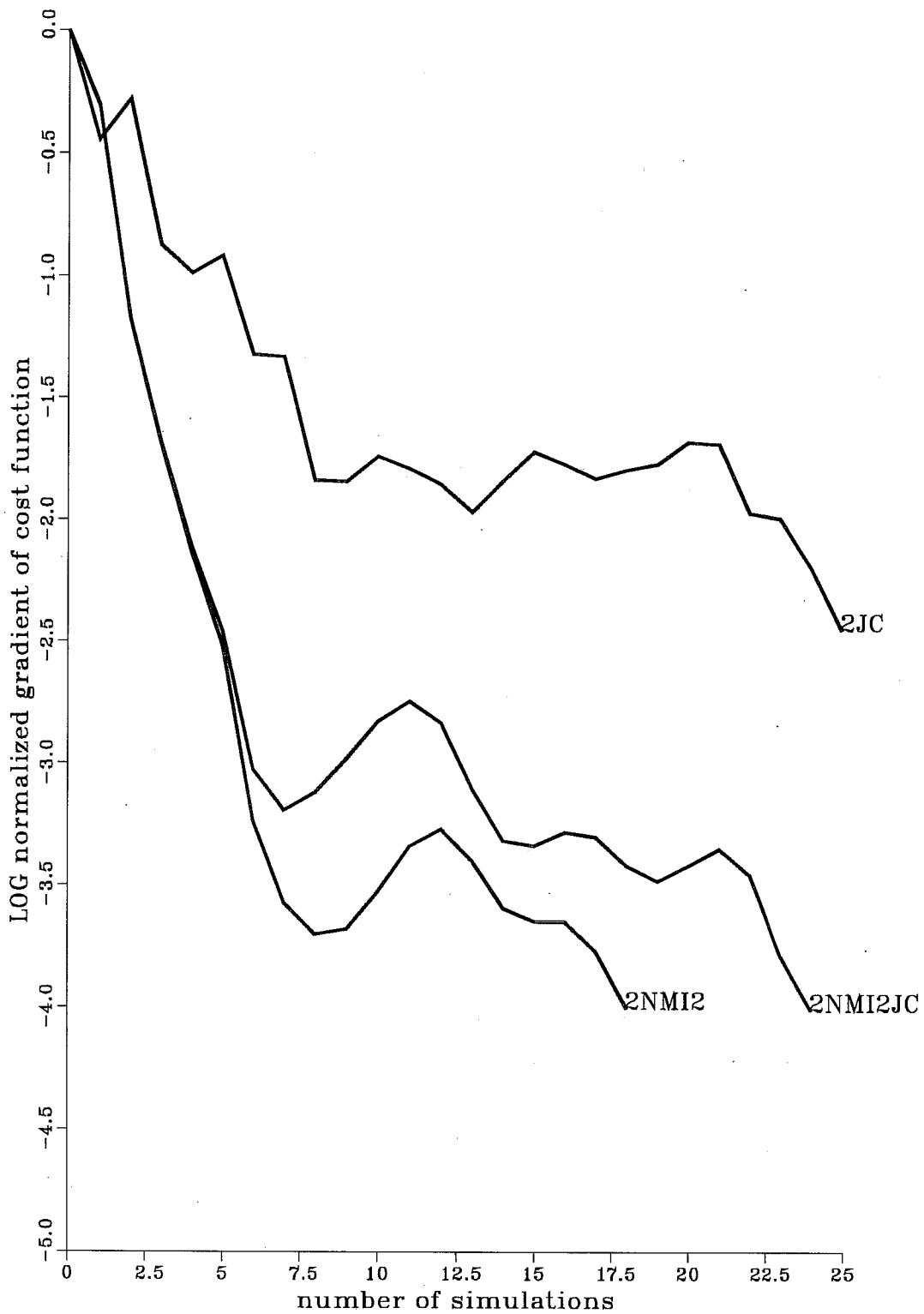
Fig. 3 Decimal logarithm of the square of the gradient norm (normalized by its initial value) for different experiments: *NMI* alone, constraint) J_c alone, and *NMI* + J_c. Note that none of these curves are strictly comparable with Figs. 7 and 8 as different initial conditions are used.
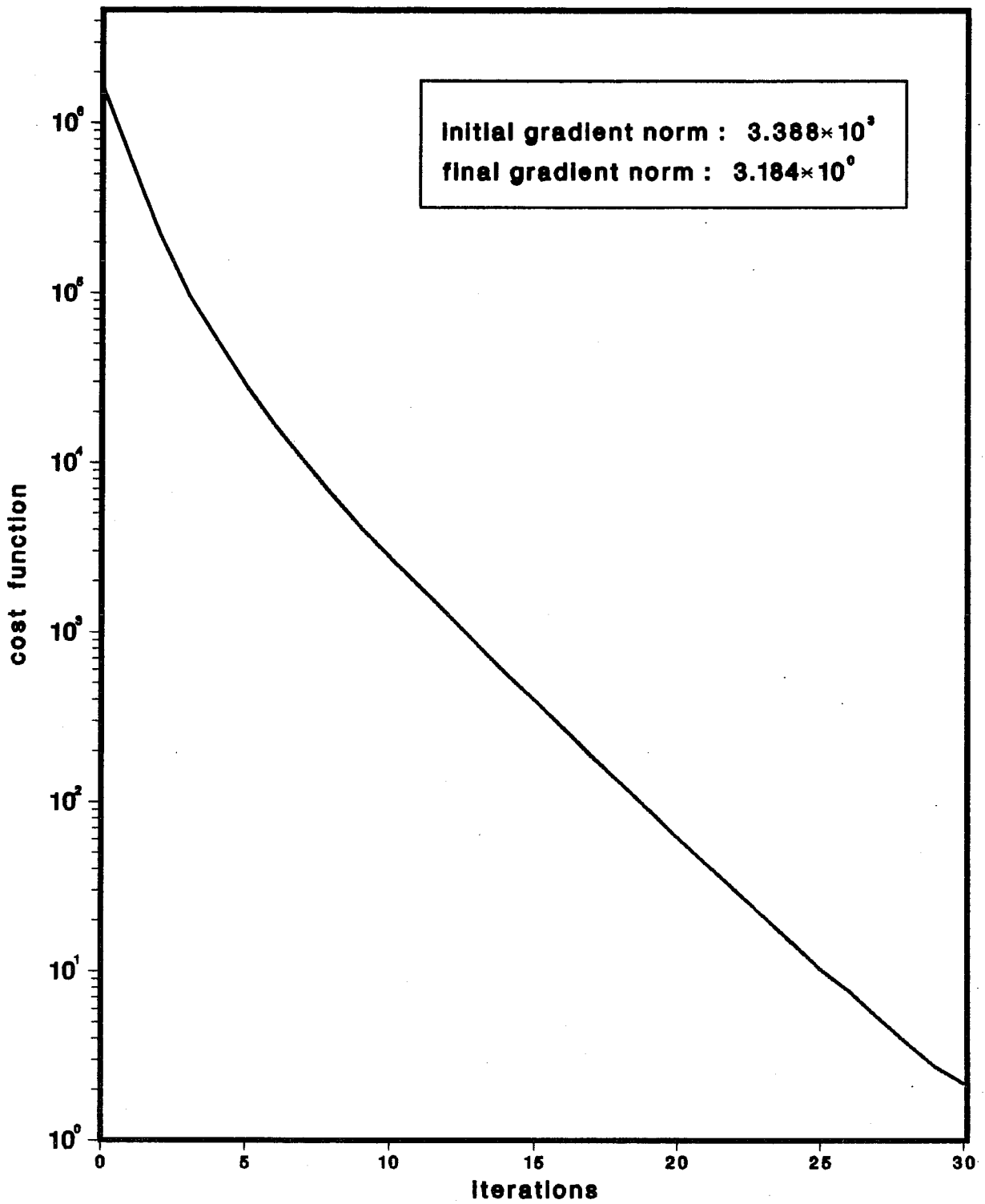
Fig. 4   Variation of the cost function with the number of iterations of the minimization process, in a 4D-assimilation experiment consisting of "inverting" the model integration on a 6 hour period.
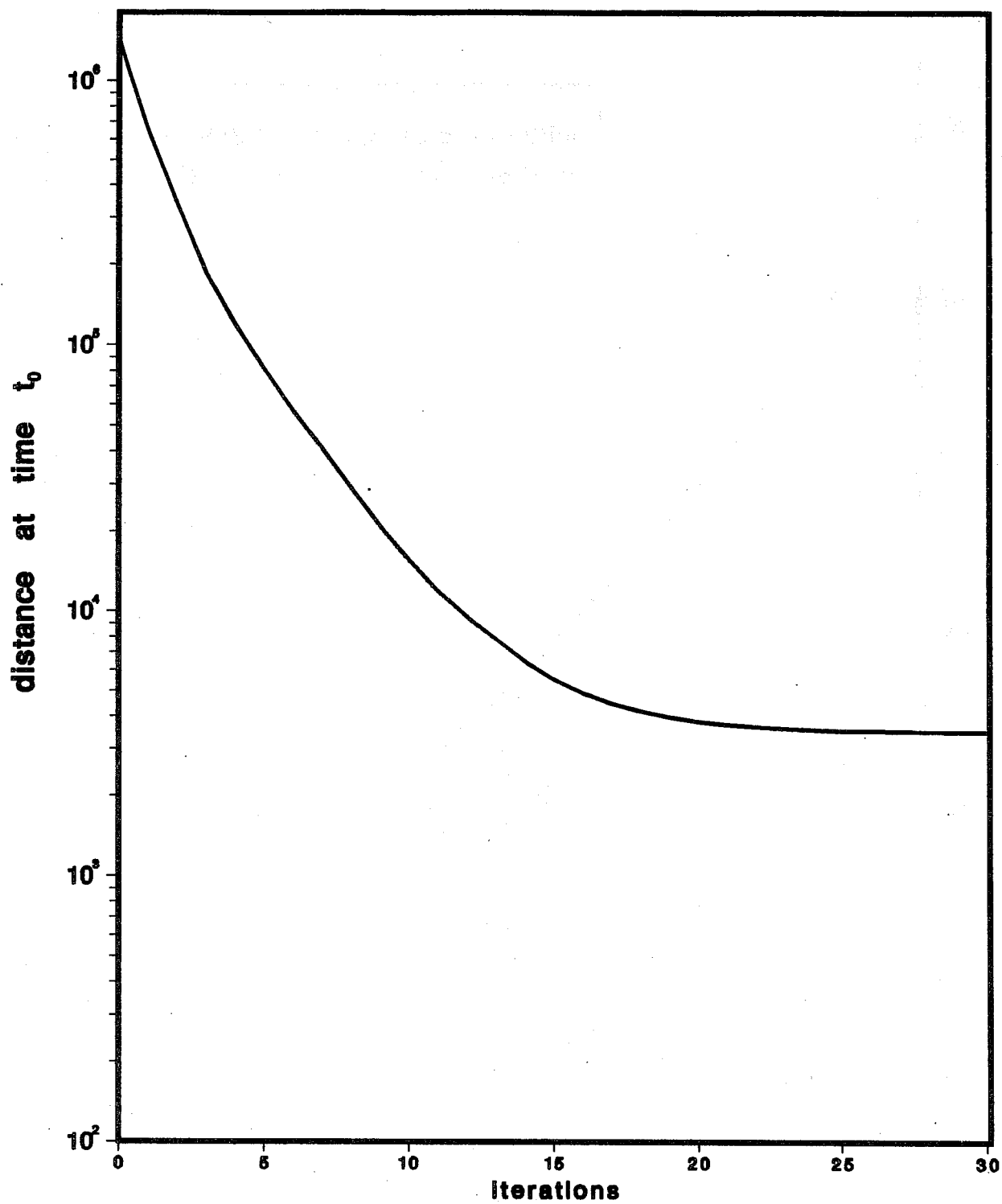
Fig. 5   Variation of the distance between the reference and the model state at the initial time of the assimilation period, during the same experiment as in Fig. 1.
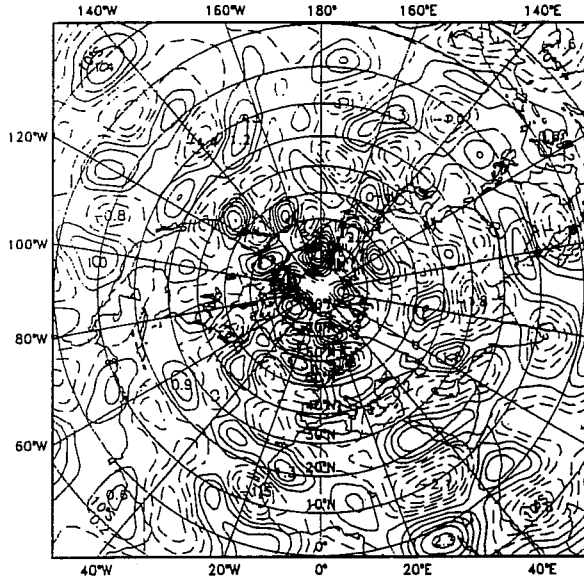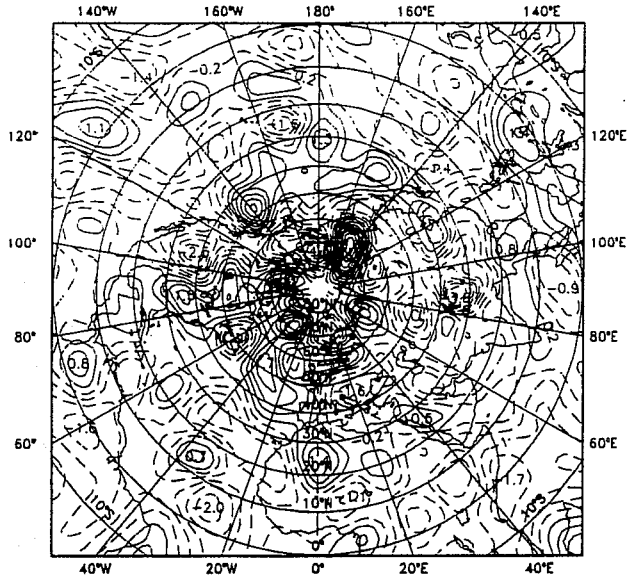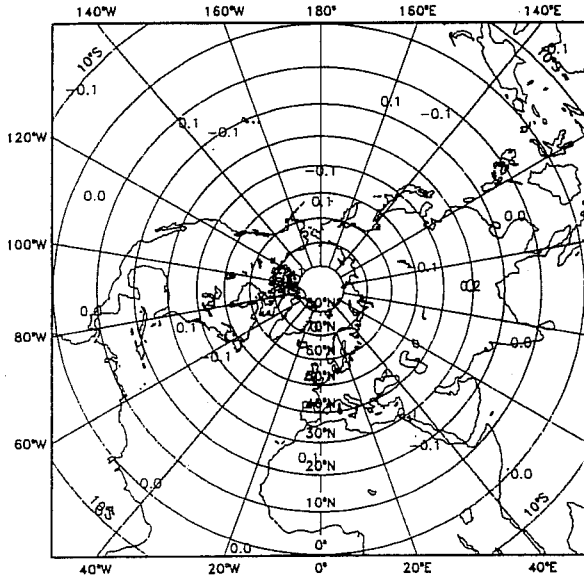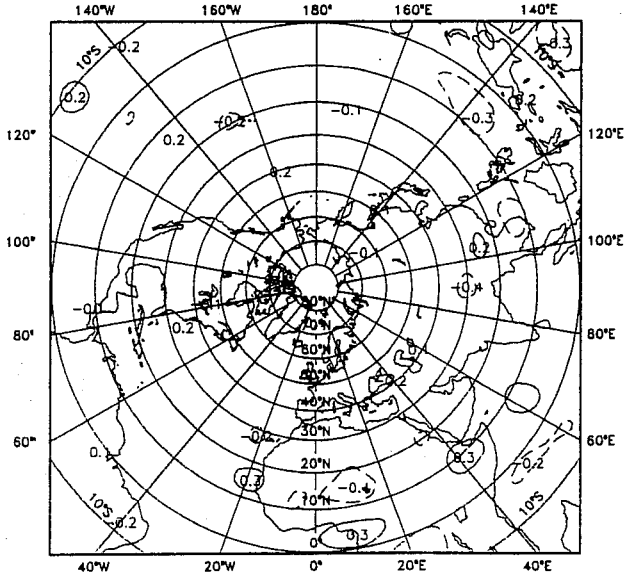
Fig. 6   Vorticity (left) and temperature (right) difference fields at 500 hPa:
Top: difference between the starting field for the minimization and the reference.
Bottom: difference between the final point after 30 iterations and the reference.

The contour interval is 0.4 10^{-5} s^{-1} for the vorticity field and 0.4°K for the temperature field.

LEV : 500 hPa   RMS = 0.832E-5 S-1
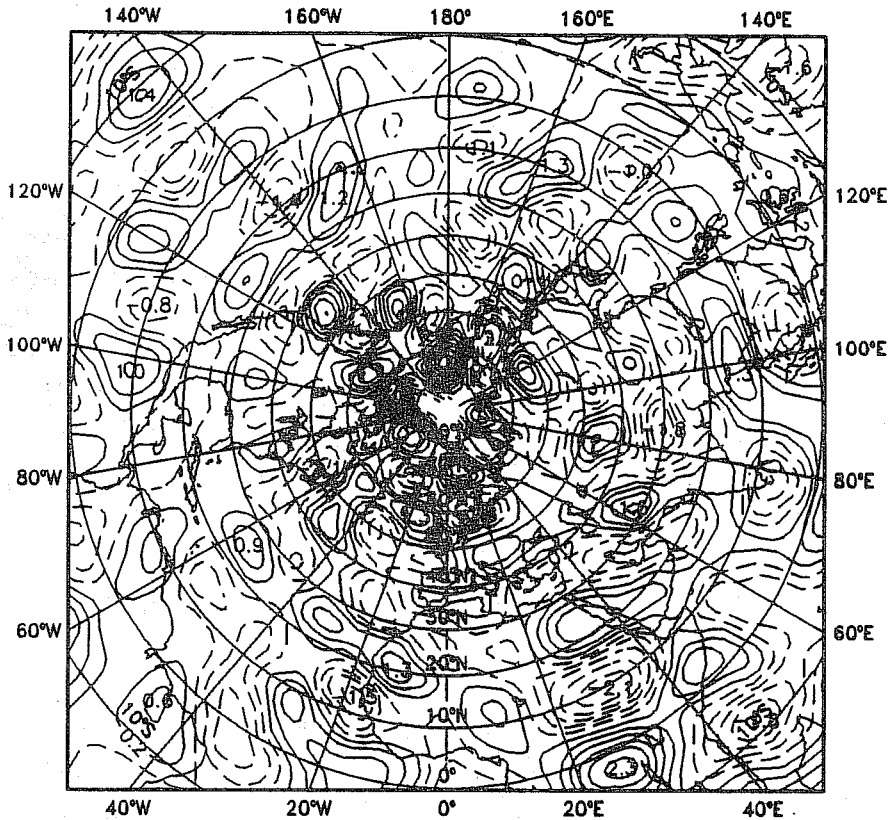
LEV : 500 hPa   RMS = 0.391E-5 S-1
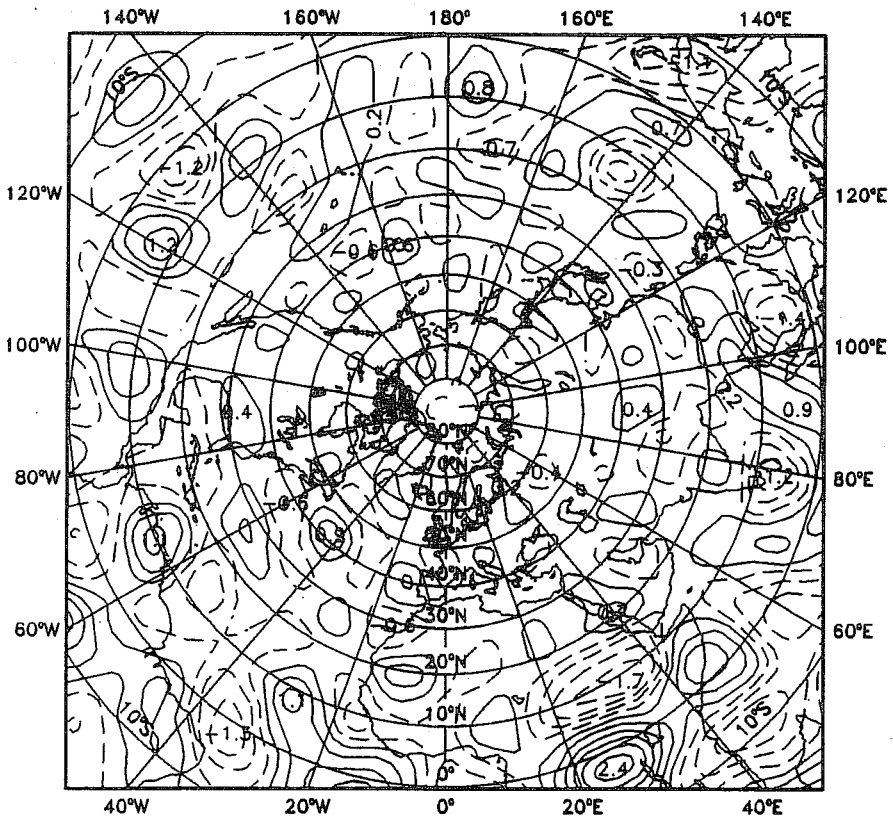
Fig. 7   Vorticity difference fields at 500 hPa before and after the minimization, when only the mass field is observed every hour.

Top: difference between the starting field for the minimization and the reference.
Bottom: difference between the final field after 30 iterations and the reference.
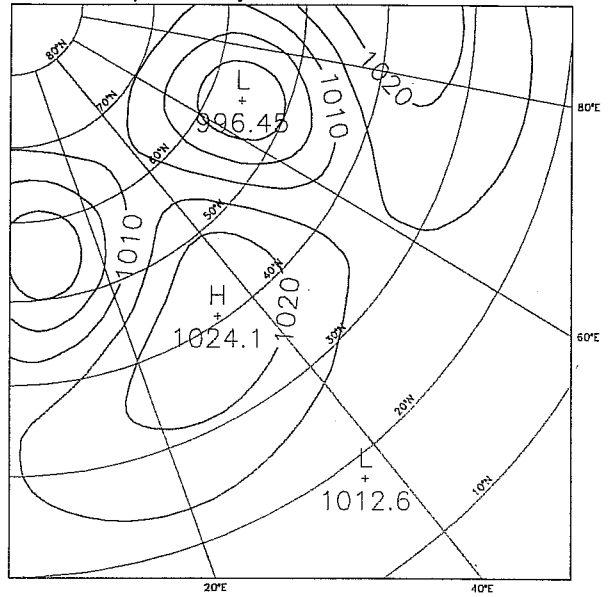
The contour interval is 0.5 10^{-5} s^{-1}.

48
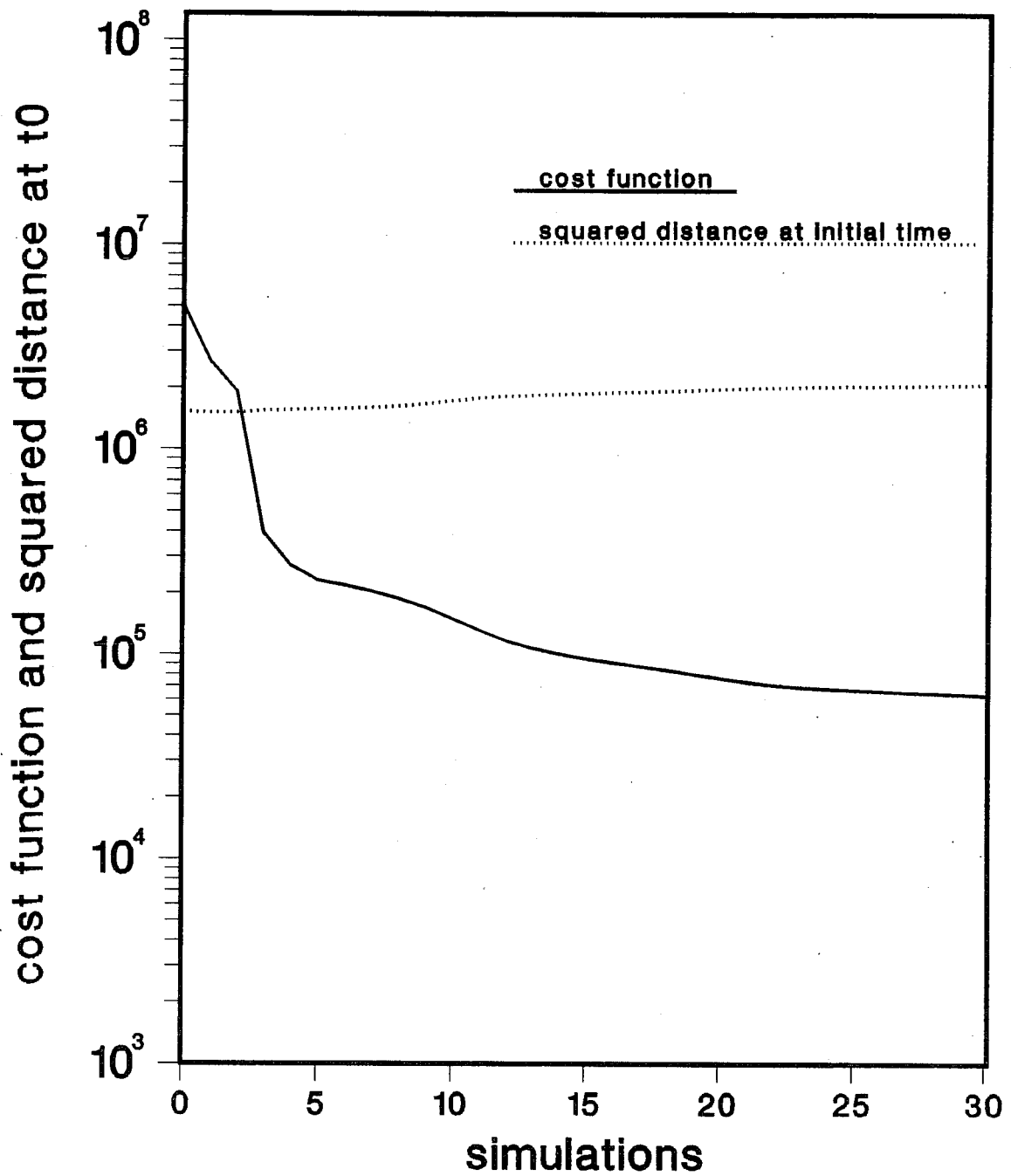
Fig. 8   Variations of the cost function and of the distance to the reference at the initial time of the assimilation period. In this 4D experiment, one tries to reconstruct a baroclinic wave from the observations of the time evolution of the rest of the flow.

49

EXP 1 surface pressure day 4 level 19
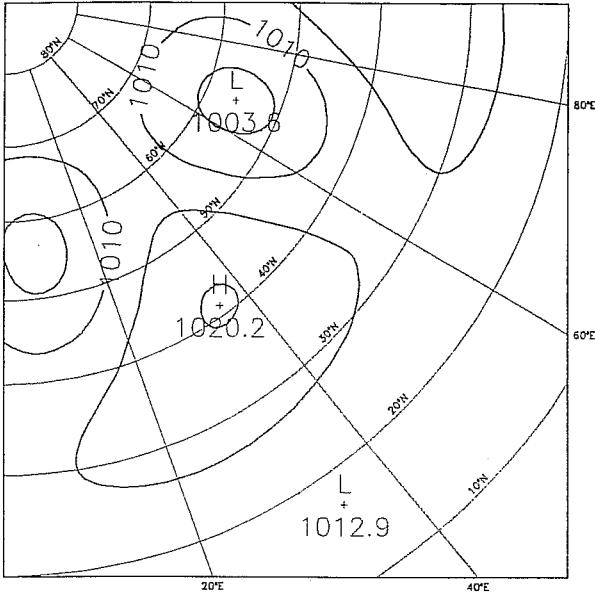
EXP 7 surface pressure day 6 level 19

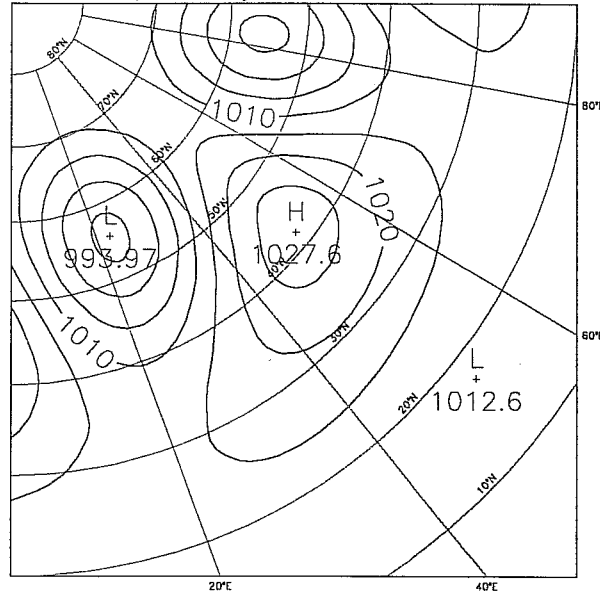EXP 1 surface pressure day 6 level 19

Fig. 9   Mean sea level pressure maps corresponding to the baroclinic wave one tries to reconstruct.
a)   initial point of the minimization.
b)   field after 30 iterations of the minimization scheme.
c)   reference field (as one would like the minimization to reconstruct it).

variation of the normalised cost function
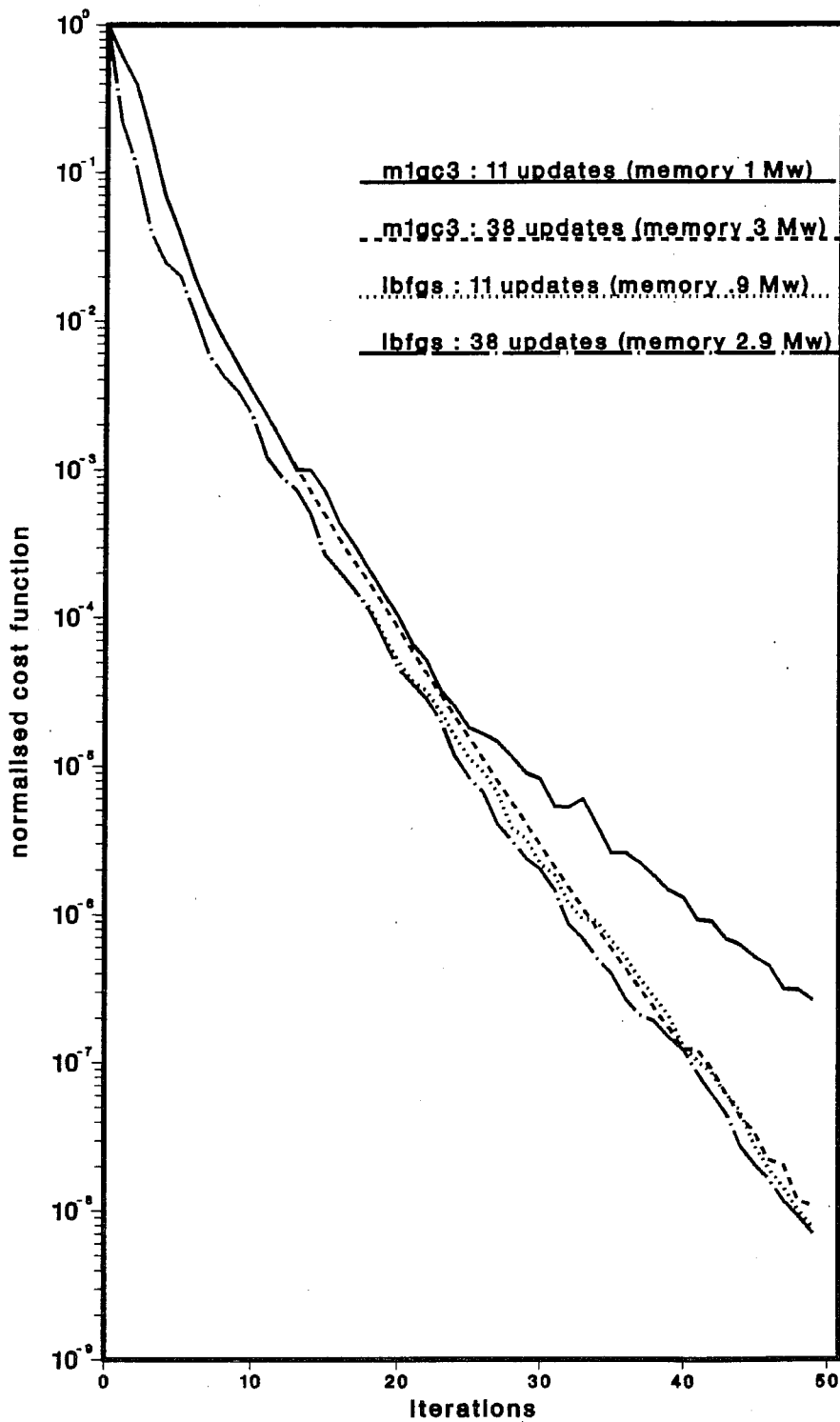
with the number of iterations

Fig. 10   Variation of the cost function normalized by its initial value with the number of simulations of the minimization scheme, comparing different versions of the minimization scheme. Note that, except M1GC3 with 11 updates, the three other versions show exactly the same performance.