TECHNICAL MEMORANDUM

178

Four dimensional variational data assimilation using the adjoint of a multilevel primitive equation model

J N. Thépaut and P. Courtier

Research Department

February 1991

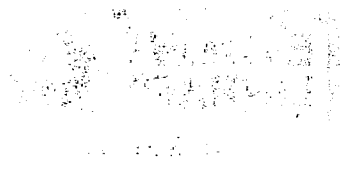
European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Abstract

The aim of the paper is to demonstrate the numerical feasibility of 4-D variational assimilation using a multilevel primitive equation model. The experiments consist in minimizing the distance between the model solution and the observations. The gradient of the cost function thus defined is computed integrating the adjoint of the model.

Here, assimilations are performed using model generated observations. In a first set of experiments, assimilations are performed assuming that observations consisting of a full model state vector are only available at the end of the assimilation period. The numerical convergence of the method is proved and the results are meteorologically realistic. The use of the Machenhauer non linear normal mode initialisation scheme and of its adjoint turns out to speed the convergence and to control to some extent the amount of gravity waves in the solution. We identify a loss of conditioning of the minimisation problem with an increase of the length of the assimilation period. The impact of the presence of horizontal diffusion in the model is also discussed.

The second set of experiments evaluates the impact of observations distributed over the whole assimilation period. Through different scenarios of sets of observations, we demonstrate the efficiency of the 4-D variational approach to extract the information contained in the dynamics of the model together with the information contained in the observations. In particular, observing only the small scales of the flow leads to a good reconstruction of both the small scales and the large scales. The observation of the mass field evolution leads to a good reconstruction of the vorticity field in the mid-latitudes and to a poorer under the tropics.



1 INTRODUCTION.

In their review of the data assimilation problems in meteorology and oceanography, Ghil and Manalotte-Rizzoli (1990) stated the already well known fact (at least since Kalman, 1960) that under the linear assumption the Kalman-Bucy filter has a least-square variational formulation and vice-versa. Furthermore, they explained how the current operational implementation of optimal interpolation can be seen as an approximation of the full KB filter.

The line of work we are following is a direct consequence of the previous remark associated with an identification of the major weaknesses of the current operational data assimilation systems that have been used in the last five years. Most of the experience accumulated by the authors comes from the ECMWF and METEO-FRANCE large scale data assimilation systems (EMERAUDE for the latter) and with the mesoscale PERIDOT operational at METEO-FRANCE. See the 1988 ECMWF seminar proceedings for a rather extensive review.

ECMWF and METEO-FRANCE concluded in parallel that one of the potential solution to the identified problems was the variational scheme either in its 3-D or 4-D formulation (3-D differs from 4-D in that the time dimension is not taken into account). A major code development project has been initiated in 1988 which led to the setting up of a cooperation between ECMWF and METEO-FRANCE for the coding of a model, an optimal interpolation, a variational assimilation (3-D and 4-D) and a Kalman filter in the same framework. The project is named IFS for "Integrated Forecasting System" at ECMWF and ARPEGE for "Action de Recherche Petite Echelle Grande Echelle" at METEO-FRANCE.

Within that cooperation a special joint effort is devoted to understanding the behaviour of variational assimilation. Courtier and Talagrand (1990) (referred to as CT90 in the following) presented results with a global shallow-water model. The main conclusion was that it was possible to perform a 4-D variational assimilation on a 24 hour period (with real observations) whose quality was similar to the quality of the operational data assimilation as judged by the RMS error of a 24h forecast, and this despite the fact that the operational assimilation has used a 15 level primitive equation model with far more observations. In the light of these results, the aim of this paper is to demonstrate the numerical feasibility of 4-D variational assimilation using a multilevel primitive equation model with 19 levels on the vertical and a spectral horizontal discretisation triangularly truncated at total wave number 21 and 42. The experiments consist in minimizing the distance between the model solution and the observations. The gradient of the cost function thus defined is computed integrating the adjoint of the model.

In section 2, we present some insight into the motivation for the development of variational assimilation. In section 3, we briefly summarize the main features of the ARPEGE/IFS code. In section 4 we present the general framework of the assimilation experiments performed.

In section 5, we perform assimilations assuming that observations consisting of a full model state vector are only available at the end of the period. Within that section several numerical issues are addressed like the conditioning of the minimisation problem and the

control of the gravity waves. The goal of the second set of experiments, presented in section 6 is to evaluate to a certain extent the impact of using observations distributed over the whole assimilation period. Through different scenarios of sets of observations, we demonstrate the efficiency of the 4-D variational approach to extract the information contained in the dynamics of the model together with the information contained in the observations.

2 THE 4-D DATA ASSIMILATION PROBLEM.

Let us introduce the extended Kalman-Bucy filter (EKB filter) applied to a dynamical system whose state vector is denoted by x in a phase space E in which an inner product $< .,. >$ is defined. In the following we just summarize the main results. For a meteorological introduction to KB filter one should refer to Ghil et al. (1981), Jazwinski (1970) for a more mathematical background and Caines (1988) for more recent results. The time evolution of x is governed by the equation :

$$\frac{dx}{dt} = \mathcal{F}(x) \tag{1}$$

which after time discretization leads to a numerical model M :

$$x(t) \longrightarrow x(t+T) = \mathcal{M}(t+T,t)x(t) \tag{2}$$

We introduce the tangent linear equation which describes to first order the time evolution of a perturbation δx in the vicinity of a trajectory $x(t)$.

$$\frac{d\delta x}{dt} = \mathcal{F}'_{x(t)} \cdot \delta x \tag{3}$$

whose discretised version, obtained by linearising (2), will lead to the tangent linear model :

$$\delta x(t) \longrightarrow \delta x(t+T) = \mathcal{R}(t+T,t)\delta x(t) \tag{4}$$

As in Talagrand and Courtier (1987) , we introduce the adjoint of the tangent linear equations :

$$-\frac{d\delta' x}{dt} = \mathcal{F}'^{*}_{x(t)} \cdot \delta' x \tag{5}$$

and the corresponding discretised adjoint model :

$$\delta' x(t+T) \longrightarrow \delta' x(t) = \mathcal{R}^*(t,t+T)\delta' x(t+T) \tag{6}$$

to which it is necessary to add the inhomogeneous forcing term in order to compute the gradient of any scalar function with respect to the initial conditions of the forecast. Introducing this notation, the EKB filter consists of two basic steps. Firstly, a forecast step in which the state x and the covariance matrix P of forecast error are transported in time :

$$x(t+T) = \mathcal{M}(t,t+T)x(t) \tag{7}$$

$$P(t+T) = \mathcal{R}(t+T,t)P(t)\mathcal{R}^*(t,t+T) + Q \tag{8}$$

Q describing the model generated errors. The "extended" character of the EKB is clearly visible on this last couple of equations where the non-linear model $\mathcal{M}$ is used to transport in time the state vector x, whereas it is the tangent linear model $\mathcal{R}$ which is used to

transport the covariance of the forecast error. The implicit assumption in (8) is the good description of the forecast error evolution by the tangent linear model. It has been shown by Lacarra and Talagrand (1988) on a channel shallow-water model that it was a reasonable assumption. This has been indirectly confirmed by Courtier and Talagrand (1987) and (1990) in spherical geometry with respectively a global vorticity equation model and a global shallow-water model. To which extent the short range forecast error evolution is governed by the tangent linear approximation is a key issue for the understanding of the behaviour of modern data assimilation scheme.

Hollingsworth and Lönnberg (1986a and 1986b) have shown that the 6 hour forecast error is comparable in magnitude to the observation errors, result confirmed by Mitchell et al. (1990). In addition, a fundamental property of the dynamical system in meteorology is the divergence of the trajectories on the attractor which means that the forecast errors are amplified in time with a typical doubling time of 1 to 2 days. As shown in appendix with a very crude analytical model those two facts are a strong indication that the term Q of (8), the model generated errors, is smaller than the first term on the right hand side of (8). This has to be understood in a mean statistical sense and has to be qualified by the fact that it is likely not to be true when the weakest part of the model is the dominant physical phenomenon like in a convective situation.

At some stage, an observation y_i available at point i related to the model state variable x by the observation operator linear or not H_i is available :

$$y_i = H_i(x) + \varepsilon_i \tag{9}$$

where ε_i, whose covariance matrix is denoted by O_i, models the observation errors (representativness errors, instrumental errors, errors in the observation operator H, see Lorenc, 1986 for a thorough discussion). Once such an observation is available, one performs the analysis step :

$$x_a = x_g + K_i(y_i - H_i(x_g)) \tag{10}$$

where x_a is the analysed state, x_g is the first guess (given for instance by a short range forecast). The expression for K_i is given by the minimum variance optimality condition :

$$K_i = P_g H_i'^t (H_i' P_g H_i'^t + O_i)^{-1} \tag{11}$$

where H_i' is the tangent linear operator of H_i and P_g the first guess error matrix (the superscript t denotes the matrix transpose). Again, one sees here the extended character of the KB filter with the full non-linear operator H_i used to compare the model to the observations in (10) and the tangent-linear operator H_i' used for deriving the weights K_i given to the observations in (11). The covariance matrix of errors of the analysis state is given by :

$$P_a = (I - K_i H_i') P_g \tag{12}$$

Using for x_g and P_g the result of the forecast step and x_a and P_a as initial conditions for the next forecast, one has the implementation of the EKB filter.

As already said, the operational implementation of optimal interpolation is an approximation of the EKB filter. The main approximations are :

i) The model is integrated for describing the temporal evolution of the atmosphere whereas the covariance matrix of estimation error P is assumed to take the form $P = Variances \times Correlations$ with the correlations constant in time, with a very crude hypothesis on their spatial structure (separation vertical/horizontal, quasi horizontal homogeneity) and with a very simple law for the temporal evolution of the variances. The validity of this assumption is often referred to as being the problem of deriving and using flow dependent structure functions in the OI literature.

ii) The observation operator is simplified. As an example for the satellite radiances data, the radiative transfer equation is weakly non linear in clear sky conditions and highly non linear with respect to cloudiness. The direct consequence is that the information content of a set of radiances is dependent on the meteorological scene, as shown e.g. by Thépaut and Moll (1990). Currently most NWP centres use SATEM messages which are retrieved temperature (and humidity) profiles. However, in order to take into account the variation of the information content with the atmospheric scene, they should be assigned a covariance matrix of error which depends also on the atmospheric state. This is not currently done in the operational analysis and would be rather tricky to implement. Another approach operational since 1986 in the PERIDOT mesoscale data assimilation is the direct use of radiances (Durand, 1985). If this approach has the advantage of using the genuine information present in the radiances, for a not too cumbersome practical implementation in OI, the operator H' used in (11) has to be linear with constant coefficients since this operator appears implicitly in the analytical model used for the correlations between the synthetic radiances and all the other observation types.

In addition there are some other probably minor simplifications like the data selection which reduces the dimension of the linear systems to be solved.

The identification of the weakness ii) was the motivation for the development of the 3-D variational analysis at ECMWF and METEO-FRANCE. Indeed, the statistical linear regression of (10) and (11) has a variational formulation which leads to the 3-D variational scheme in which the observation operator is linear. On the other hand, the 4-D variational scheme is equivalent to the full EKB filter on a limited time period $[t_0, t_n]$ and with a crude assumption on the term Q of (8), either as in CT90 introducing a temporal weighting of the observations, or neglecting it. With the latter, one obtain the following expression for the cost-function :

$$J(x(t_0)) = \sum_i (H_i(x(t_i)) - y_i)^t O_i^{-1} (H_i(x(t_i)) - y_i) \qquad (13)$$

3 THE ARPEGE/IFS CODE.

The ARPEGE/IFS code has been developed along two main principles. First, the code will be used for many different applications (temporal integration, 3-D variational analysis, 4-D variational assimilation, Kalman filtering, search for most unstable modes...), each of which will require clearly identified inputs and outputs. Second, research studies will require the use of relatively simple models, such as the vorticity equation model, while NWP development studies will require a full multilevel primitive equation model. Consequently, the code has been developed in such a way that the various applications can be used on the various models (within the limits of course of the available memory and computer time), by simply turning on or off appropriate switches.

3.1 Features of the ARPEGE/IFS models.

The models already coded are a shallow water model, a barotropic vorticity equation model and a multilevel primitive equation model on the sphere Σ with variable resolution (see Courtier and Geleyn, 1988 for the description of the variable mesh). The vertical coordinate η is the hybrid pressure/sigma coordinate introduced by Simmons and Burridge (1981). The models use the spectral representation of the fields based on the spherical harmonics with a triangular truncation (Jarraud and Simmons, 1983). The time scheme used here is an Eulerian semi-implicit leapfrog with an Asselin time filter. The option of a collocation grid with a reduction of the number of points in the vicinity of the poles has not been used in the experiments of this paper, neither the physical parametrisation package, the semi-lagrangian scheme or the stretched coordinate.

Some experiments have used the following horizontal diffusion :

- the e-folding time τ of the smallest resolved horizontal scales N is of 4 hours.
- it is proportional to the inverse of a standard atmosphere density and thus increases in the stratosphere.
- for a given wave number n, the e-folding time is $\tau F(\frac{n}{N})$ with $F(x) = (Max(0, 2x-1))^2$ which is very close to the standard ∇^4 operator but which allows no diffusion at all for the larger scales.

The adiabatic equations which govern the evolution of the wind $\vec{v}$, the temperature T and the specific humidity q_v over the sphere Σ are :

$$\frac{d\vec{v}}{dt} + \underbrace{2\Omega \times \vec{v}}_{\text{Coriolis}} + \underbrace{RT\nabla \ln p + \nabla\Phi}_{\text{pressure gradient}} = 0 \tag{14}$$

$$\frac{dT}{dt} - \underbrace{\kappa T \frac{\omega}{p}}_{\text{conversion}} = 0 \tag{15}$$

$$\frac{dq_v}{dt} = 0 \tag{16}$$

with,

$$
\begin{aligned}
R &= q_a R_a + q_v R_v \\
c_p &= q_a c_{pa} + q_v c_{pv} \\
\kappa &= \frac{R}{c_p}
\end{aligned}
\tag{17}
$$

R_a and R_v are the gas constants for the dry air and the water vapour, c_{pa} and c_{pv} the specific heat at constant pressure for dry air and water vapour. ∇ is the first order differential operator over the sphere and $\Delta = \nabla^2$ the Laplacian operator. The geopotential Φ is computed through the hydrostatic equation :

$$
\frac{\partial \Phi}{\partial \eta} = -\frac{RT}{p}\frac{\partial p}{\partial \eta} \tag{18}
$$

from the orography Φ_s which is taken in the ECMWF data bank for our experiments.

The adjoint of each individual routine has been coded for the simplest inner product, i.e. the canonical Euclidean inner product, the only exception being the spectral transforms for which we use Parseval equality which states that the transform is an isometry for the L^2 norm. The adjoint of the spectral transform is then its inverse. As already said in Courtier (1987), a change of inner product is easily handled by a matrix multiplication. This feature is implemented allowing us to change the metric of the minimisation with an automatic adaptation of the adjoint.

3.2 Numerical validation

The direct forecast model has been developed with both its tangent-linear and adjoint versions. We have then performed the following validations :

- forecast model itself (in its adiabatic version).
- tangent-linear versus direct model.
- adjoint versus tangent-linear.

All these tests configurations have been implemented in the code itself in order to be able to check any future modification of the model (direct, tangent-linear and adjoint) with only a reset of an appropriate switch, thus allowing a comfortable scientific development of the model.

3.2.1 The forecast model

The validation of the dynamics of the ARPEGE/IFS multilevel PE model has been done through a comparison with the current operational ECMWF model in its T21L19 version. With the same values for the fundamental constants (this was necessary since the values in ARPEGE/IFS are more up to date), the coefficient for the integration of the hydrostatic at the last level (1 in ARPEGE/IFS, ln(2) in the ECMWF model), and the initial conditions identical to 6 digits for packing reasons, two four-day forecasts have been performed in which at least 4 digits remained identical in the forecasted fields, the worst beeing for the divergence field. We concluded on that the code was correct.

3.2.2 The tangent-linear version.

We have checked that the tangent-linear version of the model is the linearisation of the direct model in the vicinity of a given trajectory. For a state variable x and a perturbation δx. The Taylor formula leads to :

$$\lim_{\delta x \to 0} \frac{\mathcal{M}(x + \delta x)_i - \mathcal{M}(x)_i}{(\mathcal{R} \cdot \delta x)_i} = 1 \tag{19}$$

when the denominator is different to 0 and where the subscript i denotes the i^{th} component of a vector. In addition to the value of 1 for the limit, we check that this limit is reached linearly which proves that all the linear part of the Taylor development is explained by the tangent linear model, as expected.

3.2.3 The adjoint version.

For a given inner product $<,>$, We check that the following equality holds to the accuracy of roundoff errors for given x and y vectors:

$$< \mathcal{R} \cdot x, y > = < x, \mathcal{R}^* \cdot y > \tag{20}$$

We also test the validity of the gradient of the scalar function $J(x)$ to be minimized as detailed in Courtier (1987). The Taylor formula applied in the direction $\overrightarrow{\text{grad}}\, J$ leads to :

$$\lim_{\alpha \to 0} \frac{J(x + \alpha \overrightarrow{\text{grad}}\, J) - J(x)}{< \overrightarrow{\text{grad}}\, J, \alpha \overrightarrow{\text{grad}}\, J >} = 1 \tag{21}$$

and again, we check the value of 1 for the limit and in addition that the residual is linearly reaching 0. In the previous formula, any direction could have been chosen instead of the gradient direction. However this particular choice leads to a reasonable scaling of the various components. We have insisted on the residual linearly reaching 0 and this on a wide range of magnitude of α since if the major bugs have of course been found through the value of the limit, the marginal bugs (marginal in the sense that they had no impact on the result of the minimisation) have been effectively detected with this test. This test is implemented using the same interfaces as the minimisation package allowing a control of the gradient at the beginning and at the end of the minimisation. The authors have found that it is very important to call this test each time a single line of code is modified in order not to waste time in interpreting apparent ill-conditioning of the minimisation only due to an incorrect gradient.

4 THE EXPERIMENTS.

4.1 General description.

As stated in introduction, the aim of this work is to evaluate the numerical feasibility and the scientific potentialities of a 4-D variational assimilation in which the information contained in the dynamical equations is fully and consistently used. Over a given time period $[t_0, t_n]$, the two available sources of information are a set of observations and the dynamical equations of the model. We then try to find the trajectory of the model which lies as close as possible to the observations.

We only consider here the adiabatic problem, the specific humidity variable is then not used. We assume that at the times t_i a full model state $x_{ref}(t_i)$ is available as an observation vector. (13) then becomes :

$$J(x) = \sum_{i=0}^{n} < x(t_i) - x_{ref}(t_i), x(t_i) - x_{ref}(t_i) > \tag{22}$$

which measures the misfit between the model state $x(t_i)$ and the observations $x_{ref}(t_i)$ *over the entire assimilation period.* All the experiments we perform consist in minimizing such a cost function in the space of the model state at the initial time t_0 of the assimilation period. The model state x consists of the vorticity field ζ, the divergence field D, the temperature field T and the logarithm of the surface pressure field $\ln(\pi)$. As inner product $< ., . >$ we use :

$$< x, x >= \frac{1}{2} \int_0^1 \iint_\Sigma (\nabla \Delta^{-1} \zeta \cdot \nabla \Delta^{-1} \zeta + \nabla \Delta^{-1} D \cdot \nabla \Delta^{-1} D + R_a T_r (ln\pi)^2 + \frac{C_p}{T_r} T^2) d\Sigma (\frac{\partial p}{\partial \eta}) d\eta \tag{23}$$

which is a quadratic invariant of the linearised primitive equations in the vicinity of a state of rest defined by a uniform surface pressure π_r and a constant reference temperature profile T_r (Talagrand, 1981). The choice of an energy form for the cost function defines a physical distance for comparing atmospheric fields and this natural weighting ensures a proper and reasonable scaling. From section 2, we see that the statistical interpretation of this choice is that the observations errors are supposed to have an equally distributed energy in between the different fields. A direct consequence of the invariance of a scalar product by a linear set of evolution equations is that the normal modes of this set are orthogonal for this particular scalar product. The Rossby modes and the gravity modes being the normal modes of the linearised primitive equations are then orthogonal.

This scalar product is also used for defining the metric in the minimisation space. One should note that this choice is the best in the sense that it is identical to the inner product defined by the hessian of the cost-function, firstly in the trivial case of a null length for the assimilation period and secondly in the vicinity of the state of rest used. There is no reason for this metric to remain optimal in the vicinity of a realistic atmospheric state and indeed the numerical results will show the contrary. However with no a priori knowledge on what could be optimal, this choice is very reasonable.

4.2 The "observations".

The easiest way to evaluate the numerical efficiency of a 4-D variational assimilation and the quality of convergence of the minimisation algorithm is to know the solution one wants to find and the value of the cost function at the minimum. The identical twin experiments where the "observations" are generated by the model itself provide the appropriate framework. We have then only to specify the starting point of the reference run over the assimilation period and also the initial point which will be used for starting the minimisation process. We have proceeded as follows :

- First, starting from an initialised analysis of 12 July 1989, 12 UTC taken from the ECMWF data bank (called $x_{ref}(t_{-48})$), a 48 h forecast is produced and the result $x_{ref}(t_0)$ is considered as the starting point of the reference run used as "truth". The "observations" are then produced by the model integrated from $x_{ref}(t_0)$. In the same way, starting from a 24 h ECMWF forecast valid for the same date as above 12 July 1989, 12 UTC (called $x_i(t_{-48})$) we produce a 48 h forecast which constitutes the initial point $x_i(t_0)$ of the minimisation process.

Such a generation process of the fields allows on the one hand to get rid of most of the transients due to the inconsistency between our adiabatic model and the analysis and on the other hand to start the minimisation process with a difference between the initial point of the descent and the reference state corresponding to the order of magnitude of a 24 h forecast error.

For studying the problem of gravity waves, section 5.2, we constitute $x_{ref}(t_0)$ and the subsequent "observations" as well as $x_i(t_0)$ differently using respectively the uninitialised analysis of 14 July 1989, 12 UTC and the 24 h forecast valid for the same date.

This particular date of 14 July 1989 has been chosen for historical reasons, its universality ensuring us the generality of the results.

4.3 Minimization package.

The minimisation algorithm we use is a mixed quasi-Newton/conjugate gradient type (Buckley and Lenir, 1983). Navon and Legler (1987) provide a detailed review of the minimisation packages available for the meteorological problems and a description of the Buckley and Lenir algorithm. Complementary results showed that it is one of the most efficient presently available algorithms (Navon, personal communication).

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 spare memory, it switches to a conjugate-gradient method, but preconditioned by the last approximation of the Hessian inverse, thus changing the metric as to make the cost function more spherical.

Unless stated otherwise, the minimisation process was stopped after 30 iterations which required 31 computations of the cost function and its gradient, all of them in quasi-Newton mode.

5 INVERSION OF THE MODEL.

The principle of the experiments is very simple (see Fig. 1) : over the given time interval $[t_0, t_n]$ and starting from the reference field $x_{ref}(t_0)$, one produces a reference forecast. One assumes then that the observations consist of the complete atmospheric fields (vorticity, divergence, temperature and surface pressure) at time t_n, called $x_{ref}(t_n)$. Starting the minimisation from the initial point $x_i(t_0)$, one produces a forecast at t_n and we minimise the following cost function :

$$J(x) = < x(t_n) - x_{ref}(t_n), x(t_n) - x_{ref}(t_n) > \tag{24}$$

$< ., . >$ being the inner product defined above. In other words, we perform a variational assimilation in the space of the model state at time t_0 but assuming that observations are only available at the end of the period of assimilation. Let us call $x_{end}(t_0)$ the final model state after minimisation of the cost function (basically 30 iterations).

5.1 Basic experiment

In the "basic" experiment we describe (see Fig. 1), we perform a variational inversion of the model over a 6 h period ($t_n = t_6$). No horizontal diffusion is applied in this experiment.

In Fig. 2 we present the variation of the cost function with the number of iterations in the minimisation process. A first remark is that the method numerically works. Indeed, after 30 iterations, the cost function is reduced by 6 orders of magnitude and the norm of the gradient is also reduced by 3 orders of magnitude.

More interesting for the validation of the method is the decrease of the difference $x(t_0) - x_{ref}(t_0)$ during the minimisation process since it measures the ability of the variational inversion to recover the reference field. Fig. 3 presents the variation of the distance $< x(t_0) - x_{ref}(t_0), x(t_0) - x_{ref}(t_0) >$ with the number of iterations. The decrease is almost of 3 orders of magnitude.

It should be pointed out that due to the choice we made for the inner product, this distance is the energy of the difference between the two meteorological flows. This energy can even be split in its different contributions, rotational kinetic energy, divergent kinetic energy, potential energy with its temperature and surface pressure contribution.

Worth noticing is that the convergence as seen from Fig. 3 is not fully achieved with 30 iterations whereas the cost function is reaching 0. The dynamics of the model is the reason for the loss of conditioning of the problem. As stated in section 4.1 the metric has no reason to be optimal for the cost function considered. The consequence is that the shape of the cost function can become strongly elliptic with respect to the metric used and the gradient can even become almost orthogonal to the direction of the minimum. This explains the difference in reduction for this distance at time t_0 (3 orders of magnitude) compared to the better reduction (6 orders of magnitude) observed at time t_6 for the cost function.

However, the decrease of the distance is meteorologically acceptable as we can see from Fig. 4. This figure presents the difference between the reference and the analysed vorticity (left) and temperature (right) fields for the northern hemisphere at 500 hPa.

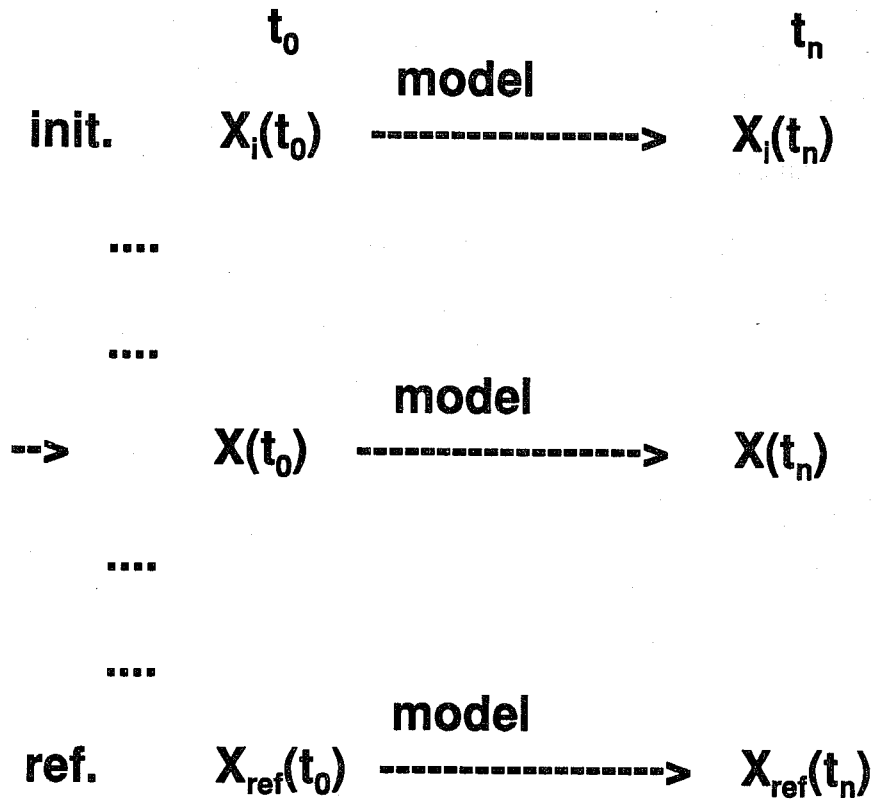


Fig. 1 Inversion of the model. The "model-generated" observations are available only at the end of the period of assimilation.

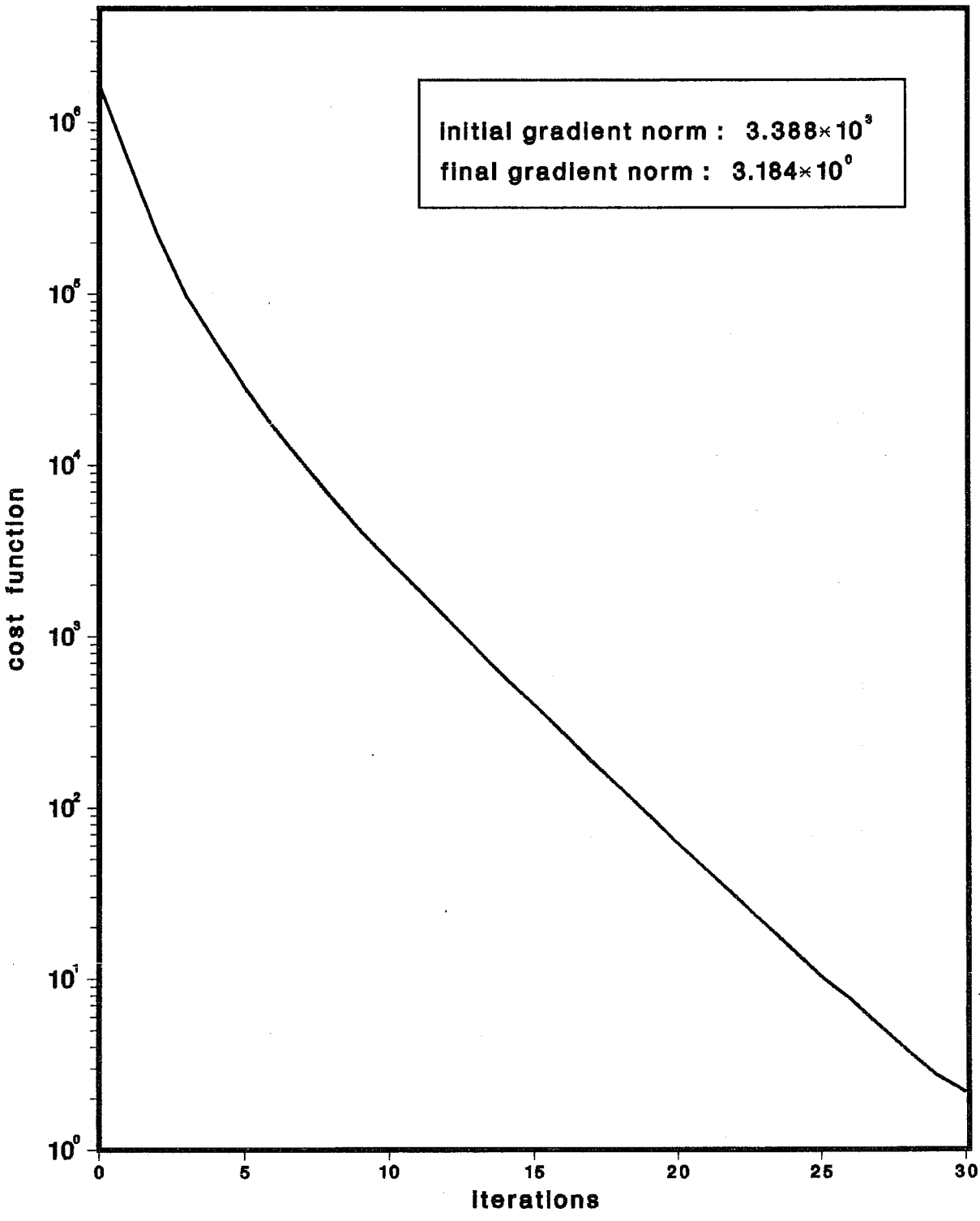


Fig. 2 Variation of the cost function with the number of iterations of the minimisation process.

Top panel shows the difference before minimisation and bottom the same difference after 30 iterations. The RMS of the differences with the "truth" are significantly reduced both for vorticity and temperature. Moreover, the differences are reduced homogeneously all over the sphere and for all the fields.

The reconstruction of the fields is also efficient regardless of the model levels and of the scales as we can see from respectively Fig. 5 and Fig. 6. Fig. 5 represents the contribution to the energy of the differences at each model level before minimisation (full curve) and after 30 iterations (dashed curve and here not lisible) for the vorticity (top panel) and the divergence (bottom panel) in J/kg. Fig. 6 represents the same distance but with respect to the total wave number n.

This experiment validates the numerical feasibility of a 6 h variational assimilation in the most difficult context where information is available only 6 hours later than the time at which the initial conditions are to be reconstructed.

5.2 Gravity waves control.

In the previous experiment no attention was paid to the amount of gravity waves contained in the solution. If the observations contain gravity waves, the result of the minimisation will also contain gravity waves. We shall now tackle the problem of finding a model trajectory which best fits the observations *and* lies on the slow manifold.

Following the idea of CT90, we run the same experiment as the basic one but with a non-linear normal mode initialisation NNMI (see Machenhauer, 1977) applied before the integration of the model both for the construction of the observations and during the course of the minimisation for comparing the trajectory of the model with the observations (see Fig. 7). 5 iterations of the Machenhauer scheme on the first five vertical modes are performed to ensure that the initialisation is close to a genuine projection on the slow manifold. We denote by $x'(t_0)$ the initialised field. As described in CT90, the adjoint of NNMI is applied at the end of the adjoint integration.

Fig. 8 presents in full line the variation of the distance between the uninitialised fields $< x(t_0) - x_{ref}(t_0), x(t_0) - x_{ref}(t_0) >$ with the number of iterations and in dashed line the distance between the initialised fields $< x'(t_0) - x'_{ref}(t_0), x'(t_0) - x'_{ref}(t_0) >$. The latter measures a distance on the slow manifold so it represents the ability to recover the Rossby part of the reference field.

The distance between the uninitialised fields decreases by less than 2 orders of magnitude. However, as stated above, a NNMI is included in the assimilation process. The NNMI is very close to a projection operator parallel to the gravity modes and then it is not invertible along the gravity modes direction (in fact, NNMI since it is iterative is only a contraction in phase space and as such the invertibility is ill conditioned). In other words it is impossible to recover satisfactorily with the 30 iterations of the minimisation algorithm performed the gravity modes information lost in the NNMI. The full line curve is then not intrinsic and the level of saturation depends on the amount of gravity waves present in x_{ref}. This is confirmed by the dashed curve which shows that the Rossby part of the flow is well reconstructed and by almost 3 orders of magnitude which is comparable to the basic experiment. The inversion of the model under the constraint that the trajectory

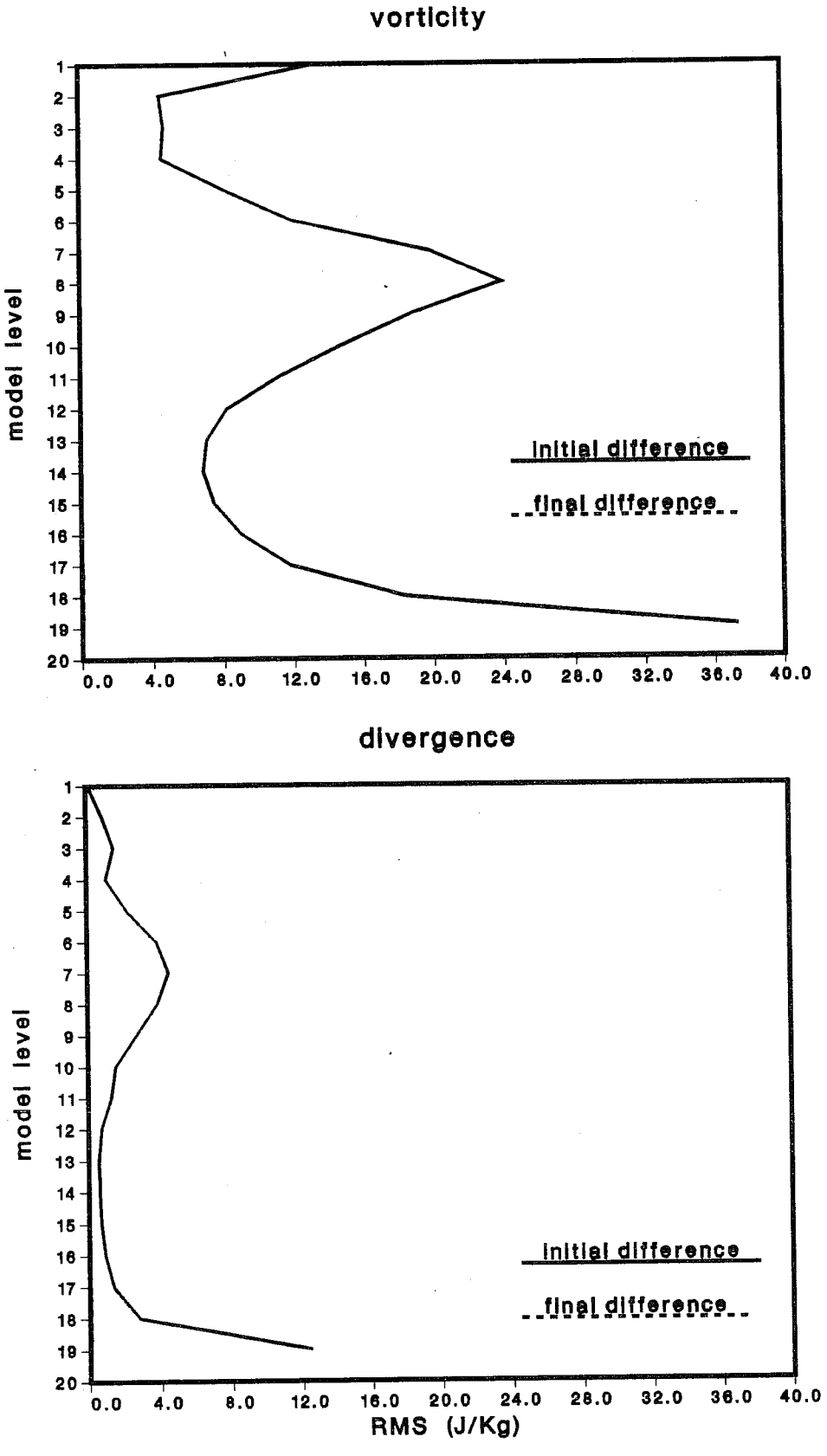


Fig. 5 Distance (energy in J/Kg) between the model state and the reference state with respect to the model levels for the vorticity field (above) and the divergence field (below).
(—) : before minimisation.
(.....) : after 30 iterations.
Note that the dashed line and the 0-axis are disconcerted. The field is then "completely" retrieved.

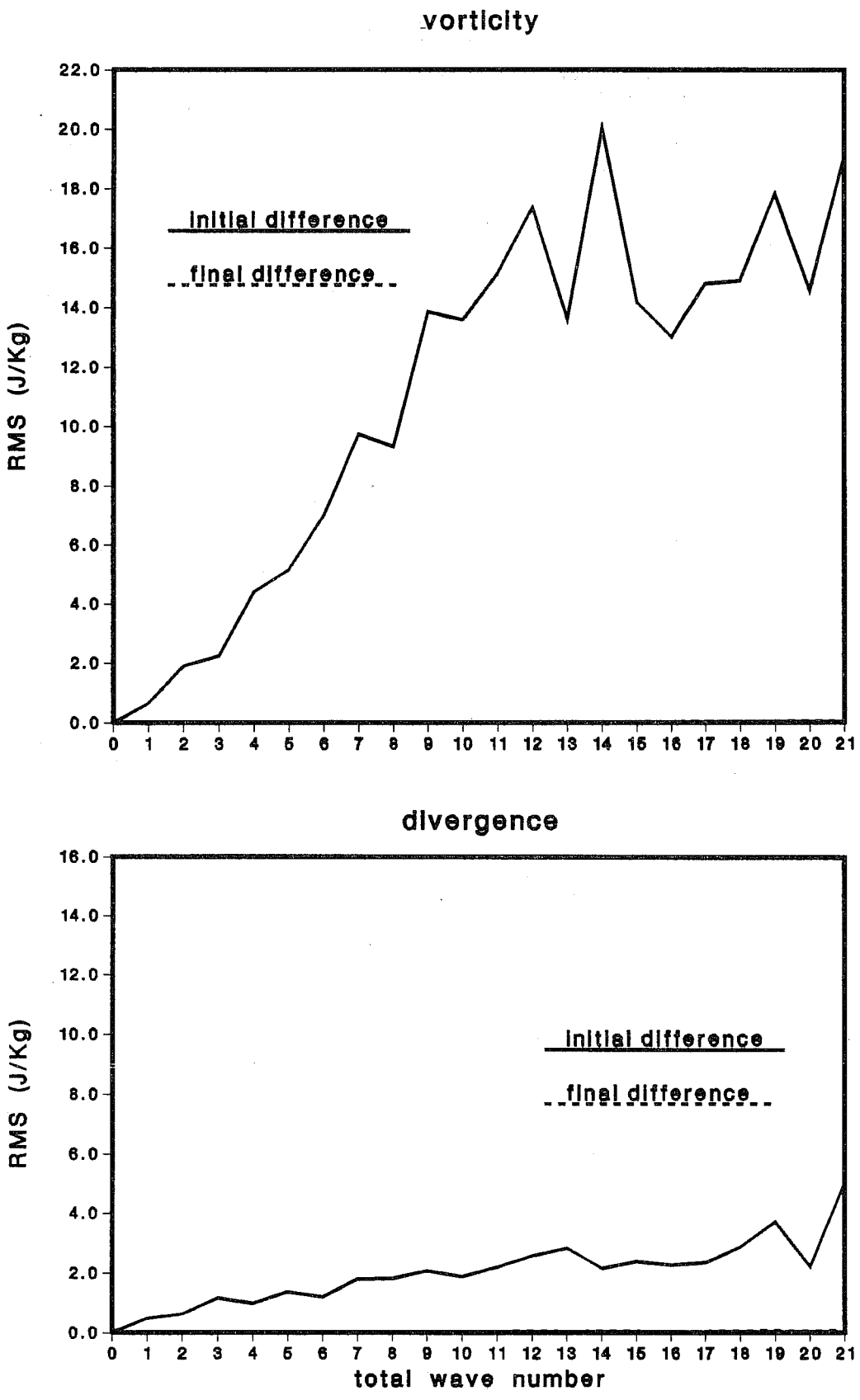


Fig. 6 Distance (energy in J/Kg) between the model state and the reference state with the respect to the total wave number for the vorticity field (above) and the divergence field (below).
(—) : before minimisation.
(.....) : after 30 iterations.
Same remark as for Fig. 5 concerning vorticity.

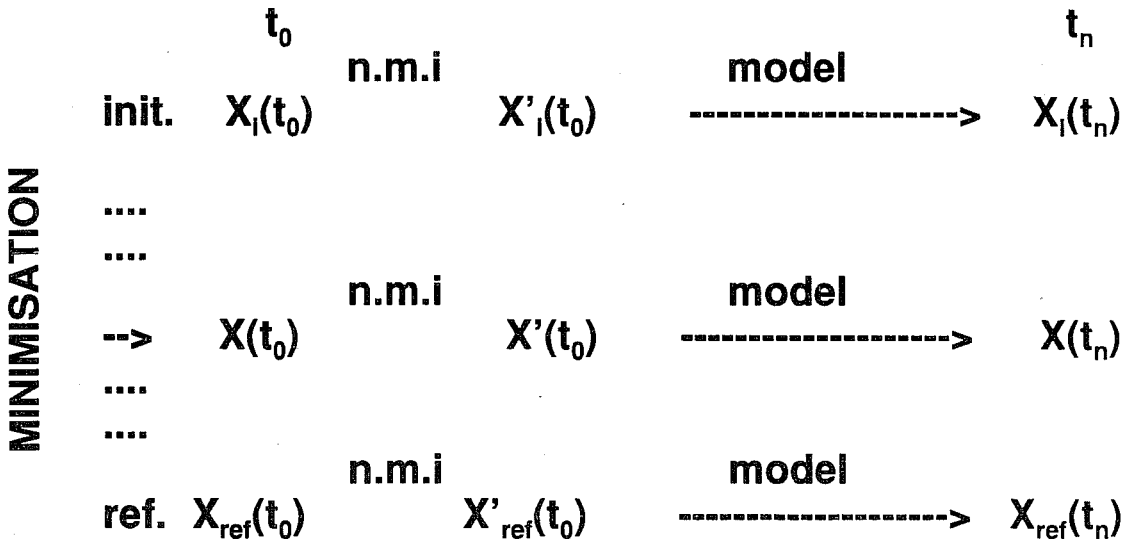


Fig. 7 Inversion of the model. A NNMI scheme precedes now the direct integration of the model.

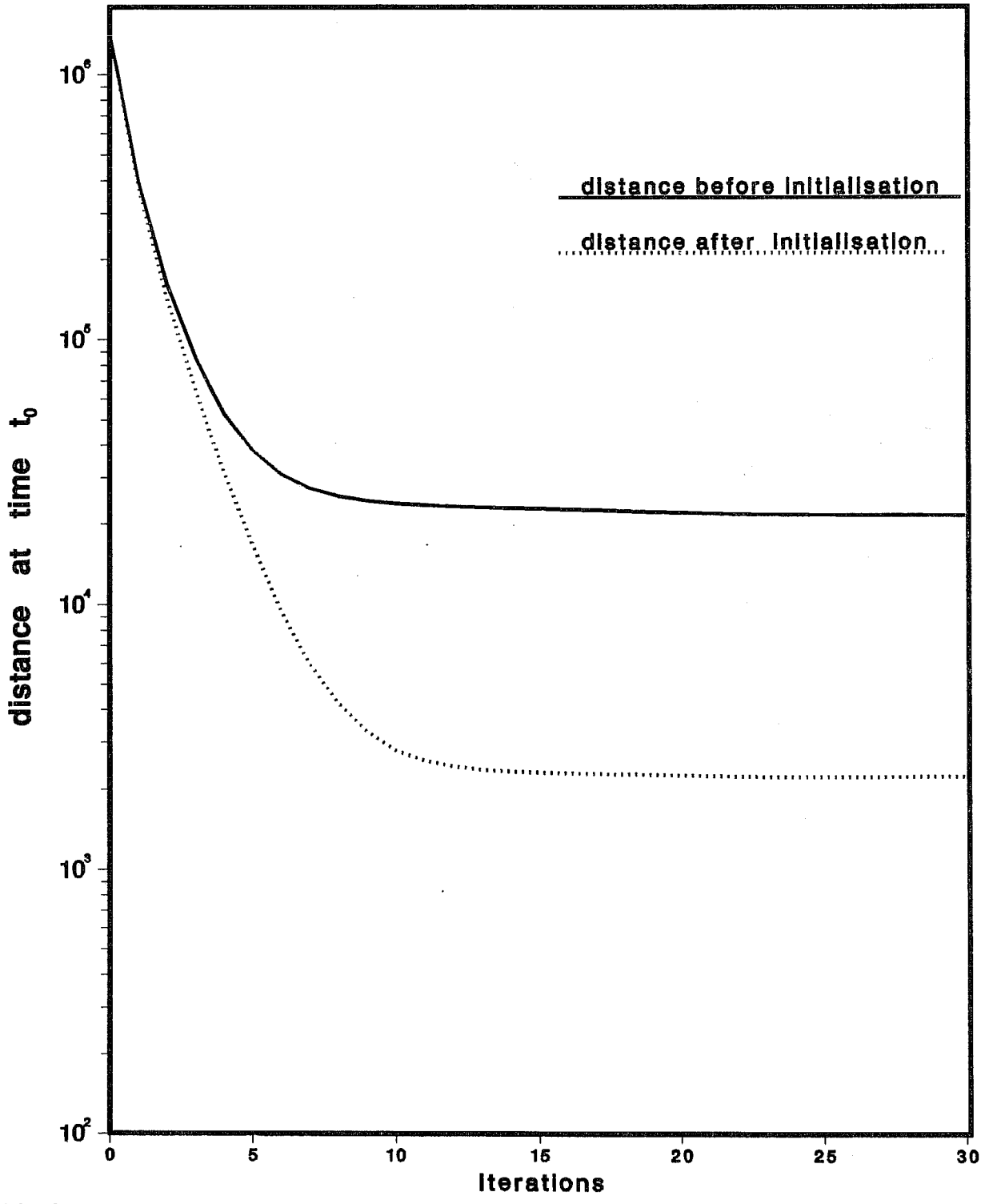


Fig. 8 Variation of the distance between the reference and the model state at the initial time of the assimilation period with the number of iterations.
(—) : distance before NNMI.
(.....) : distance after NNMI.

lies on the slow manifold is then as easy to realize as without the constraint. Moreover, the comparison of Fig. 3 with the dashed line of Fig. 8 shows that the rate of convergence is better when using the constraint, especially at the beginning of the minimisation.

However, CT90 have faced a difficulty to impose as a strong constraint the solution to lie on the slow manifold because of partial invertibility of NNMI. In order to evaluate this problem in the multilevel context, we run the same experiment as the previous one but with the reference and the initial point of the minimisation built as explained in second part of section 4.2 : $x_{ref}(t_0)$ is the uninitialised analysis of 14 July 1989,12 UTC and $x_i(t_0)$ the 24 h forecast valid for the same date. As in the previous experiment, we include the Machenhauer process before integration of the model at each step of the minimisation algorithm.

As already said NNMI is almost a projection of the entire phase space (point A) onto the slow manifold S (point B) along a direction parallel to the gravity modes subspace, and can be schematically represented by the now standard diagram introduced by Leith (1980) (see Fig. 9). A state vector x is here represented by its gravity part G and its rossby part R. The tangent linear of NNMI is then a projection onto the subspace S' tangent to the slow manifold at point B, along a direction parallel to the gravity mode subspace. And, by definition, the adjoint of NNMI projects the gradient of J on the orthogonal of the gravity modes, i.e. the Rossby modes, along a direction parallel to the orthogonal of S'. As a consequence, the gradient of the cost function with respect to the gravity components is 0.

However, as emphasized in CT90, in its implementation NNMI is an iterative algorithm stopped after a finite number of iterations, NNMI is then only a contraction in phase space and remains invertible. The condition $\| \frac{dG}{dt} \| = 0$ cannot in practice be exactly enforced and the minimum of the cost function remains theoretically unchanged. However, the retrieval of the gravity part of the flow is a badly conditioned problem, the ill conditioning increasing with the number of iterations of NNMI. Our purpose is here to evaluate this invertibility.

We present in Fig. 10 the variation of the distance between the analysed state and the reference state at time t_0 when no NNMI is applied before direct integration of the model (full line) and when the integration of the model is preceded by five iterations of the Machenhauer process (dashed line).

We can see that when we do not include the NNMI process the distance decreases significantly. The minimisation method can effectively use all the degrees of freedom of the problem, including the gravity modes. The solution matches thus the reference situation perfectly well. Our reference being an unitialised analysis, our solution contains thus a large quantity of gravity waves.

If we assume that the NNMI performed before the integration is a projection, the minimum changes since it is located on the slow manifold and only the Rossby part of the field is reconstructed (see section 5.1). Confirming the result obtained in the first experiment of this section, using a NNMI scheme speeds up the convergence. The dimension of the slow manifold (as defined by $\| \frac{dG}{dt} \| = 0$) is 23232 and smaller than the dimension of the space (28072 at truncation 21). This explains the fact that at the very beginning of the minimisation, the distance decreases faster when performing NNMI

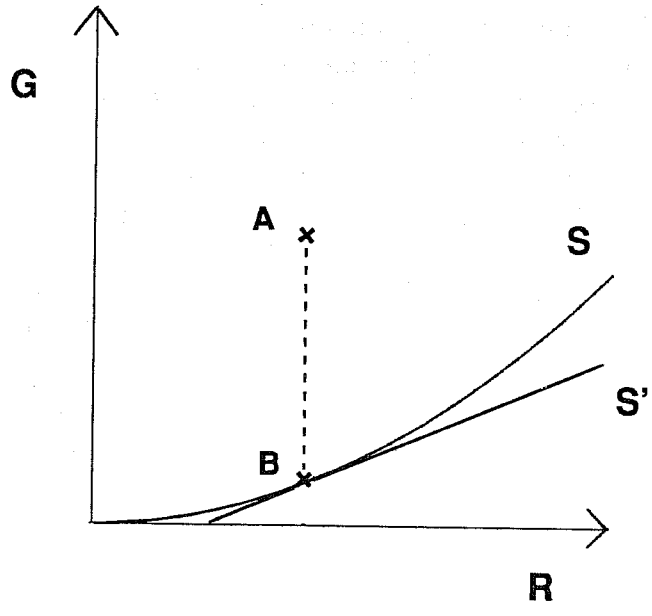


Fig. 9 Schematic representation of the normal mode initialisation (from Leith, 1980 and Courtier and Talagrand, 1990).
R and G represent the two subspaces respectively associated to the Rossby and the Gravity modes. Considering that the Machenhauer algorithm is a projection, a point A will be projected through this algorithm on the slow manifold S (point B). S' is the tangent of S at point B.

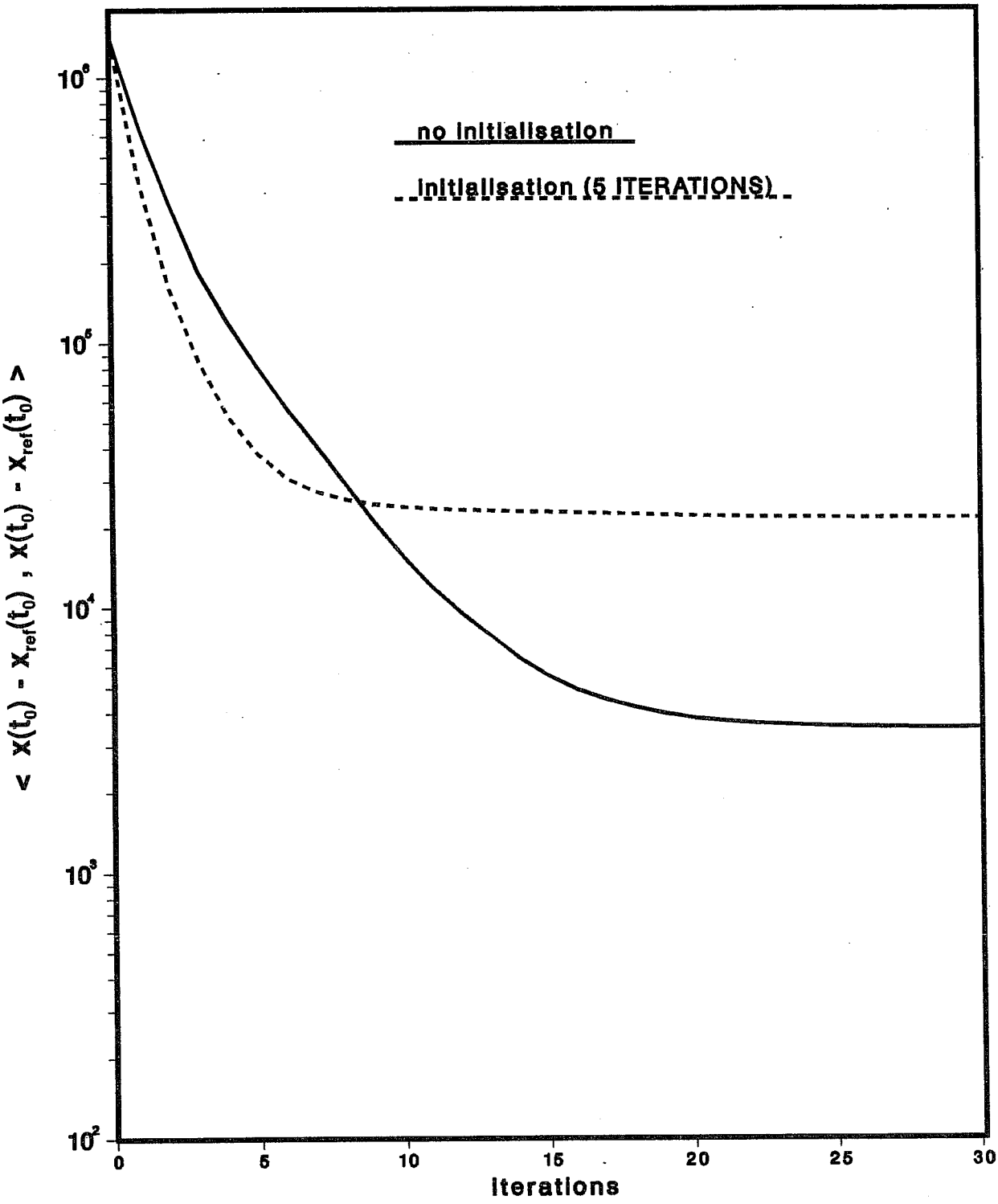


Fig. 10 Variation of the distance between the reference (here an unitialised analysis of the 14 July 1989) and the model state at the initial time of the assimilation period with the number of iterations.
(—) : no NNMI is used.
(.....) : NNMI is used before integration (5 iterations).

before integration.

Another insight on the efficiency of this constraint is to look at the term $\| \frac{dG}{dt} \|$ of the analysed state during the minimisation process (see Fig. 11). In full line we represent the energy of the gravity tendency of the reference state at time t_0 and for the first vertical mode. This energy will be denoted by BAL (for degree of balance) in the following. The other curves represent the variations of BAL for the model solutions during the minimisation process for different numbers of iterations in NNMI. Starting the minimisation from a 24 h forecast, BAL is only 25% of the value in the reference state.

On the curve corresponding to 1 iteration in NNMI, BAL has almost reached the value in the reference after 3 minimisation steps. In that case, NNMI is practically invertible. Increasing slightly the number of iterations seems only to delay the step at which the minimisation process starts the reconstruction of the gravity part of the reference state. This is only when 10 iterations of NNMI are applied that it appears not to be invertible (as far as the minimisation algorithm is concerned). This result consistent with what was found in a shallow-water model by CT90 leads to two main remarks :

- Firstly, we can expect that when dealing with real observations and starting the minimisation process from a background field free of gravity waves, the use of NNMI at each step of the minimisation will to some extent force the solution to lie close to the slow manifold, providing the ratio of the number of NNMI iterations to the number of minimisation descent steps is large enough. However, mainly for cost reasons, it seems unrealistic to use more than 3 iterations in a NNMI scheme. As in CT90, one will have to consider combining this approach with the use of a weak constraint to penalize the distance to the slow manifold. The weak constraint will change the minimum and the introduction of NNMI will speed up the convergence.

- Secondly, if we restrict the model space to two dimensions (R,G), R being the Rossby component and G the gravity component, NNMI considered as a contraction admits as eigen values respectively 1 and ϵ. ϵ gets smaller as the number of iterations of the NNMI increases, from the evolution of BAL during NNMI it typically loses one order of magnitude at each iteration. This gives us an indication on the number of iterations necessary for the minimisation to retrieve the information corresponding to the eigenvectors of the Hessian matrix of a given eigenvalue. If 3 iterations are necessary for 1 order of magnitude, 18 are necessary for 5 orders of magnitude, 25 for 8 and 30 are not sufficient for 10. The statistical interpretation of this result is that a dynamic range of 5 orders of magnitude in the variances of analysis can be reached in 18 iterations without any work on improving the conditioning of the minimisation problem. However such work is necessary either to reduce the number of iterations or improve the dynamic range.

5.3 Impact of the length of the assimilation time interval.

All the experiments performed now are using 5 iterations of NNMI since the rate of convergence obtained was better with NNMI, 5 iterations of the Machenhauer process not leading to any major inversion problem. The observations are build like for the basic experiment but with NNMI applied before the integration of the model on $[t_0, t_n]$ (see Fig. 7).

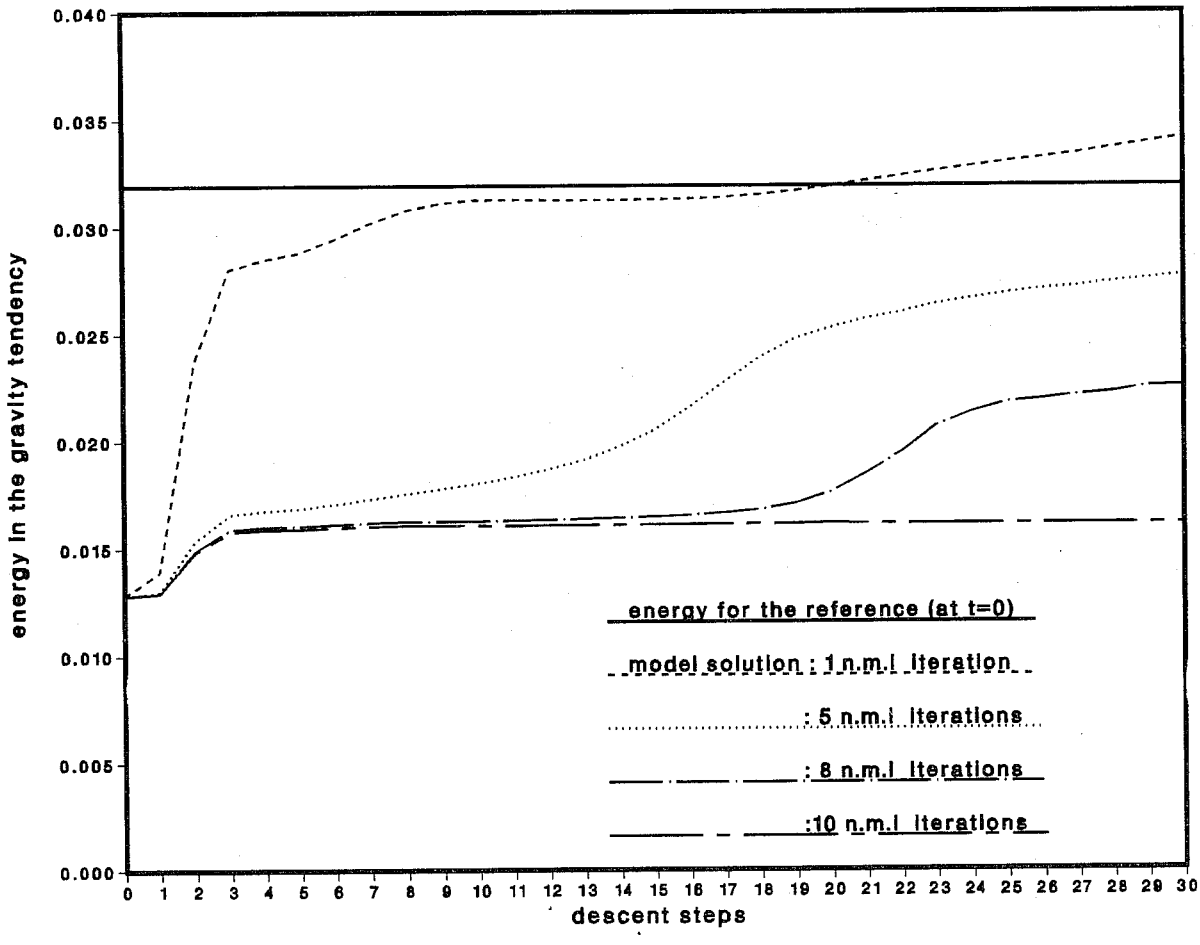


Fig. 11 Variation of the energy in the gravity tendency (BAL) for the first vertical mode of the field (at time t_o).

(————) : reference field.

(----------) : model solution (1 NNMI iteration is used).

(.....) : model solution (5 NNMI iterations are used).

(—·—·—) : model solution (8 NNMI iterations are used).

(— — — —) : model solution (10 NNMI iterations are used).

Producing the experiment with observations available at 3, 6, 12 or 18 h, we studied the impact of the length of the assimilation on the conditioning of the minimisation problem. Fig. 12 shows the variation of the distance $< x(t_0) - x_{ref}(t_0), x(t_0) - x_{ref}(t_0) >$ with the number of iterations for variational inversion periods of 3, 6, 12 and 18 hours respectively. We are looking here at the distance before initialisation. We have seen in the previous section that it was arbitrarily dependent on the amount of gravity waves present in $x_{ref}(t_0)$ but the construction of $x_{ref}(t_0)$ (result of a 48 hour model integration) ensures it to be close to the slow manifold.

In all cases, we observe that an acceptable solution is reached in 30 iterations. However, when the interval between the time when "observations" are available and the time at which the reference state has to be reconstructed increases, the rate of convergence decreases. As stated previously, the shape of the iso-J surfaces is crucial for the efficiency of the minimisation algorithm, the main difficulties of convergence being linked to the fact that the condition number of the Hessian matrix (ratio between the greatest and the smallest eigen values) becomes large. Indeed, as long as the time length of the assimilation increases, the iso-J surfaces become more and more elongated in certain directions and contracted in some other, depending on the directions in which the dynamics would amplify or contract a perturbation.

Fig. 12 shows an interesting behaviour for the 3 h and 6h inversion experiments. We observe that the convergence is fast at the beginning, but after some descent steps, the distance at time t_0 starts increasing while the cost function continues decreasing (not shown). It is in particular obvious for the 3 h experiment where the phenomenon occurs after 12 iterations.

This effect can be explained by the schematic Fig. 13. The iso-J curves on Fig. 13 are elliptic and during the minimisation, the point in phase space can move from A to B. The cost function J has decreased from A to B, but the distance to the reference at time t_0 has increased. In the course of a minimisation, such a behaviour appears generally at the separation between two clusters of eigenvalues of different order of magnitude.

In the case of Fig. 12, the ellipticity of J can be due either to the dynamics or to the NNMI. It is likely that the latter is the explanation for two reasons, firstly we do not observe this behaviour when no NNMI is applied and secondly because we have seen in section 5.2 that the minimisation algorithm starts to invert the NNMI algorithm at about this number of iterations of minimisation. NNMI has effectively clustered the eigenvalues of the Hessian in two groups, one for the Rossby modes and one for the gravity modes. The reason why we do not observe the same behaviour for the 12 h and 18 h cases could be that the dynamics has then spread the eigenvalues enough so as not to have such a clear separation between the two clusters.

5.4 Horizontal diffusion impact.

We have mentioned in section 3 that the model is used in its adiabatic version without any physics. Only a horizontal diffusion of vorticity, divergence, and temperature is available. We describe now the impact of this dissipation on the ability of the variational inversion to recover the initial conditions.

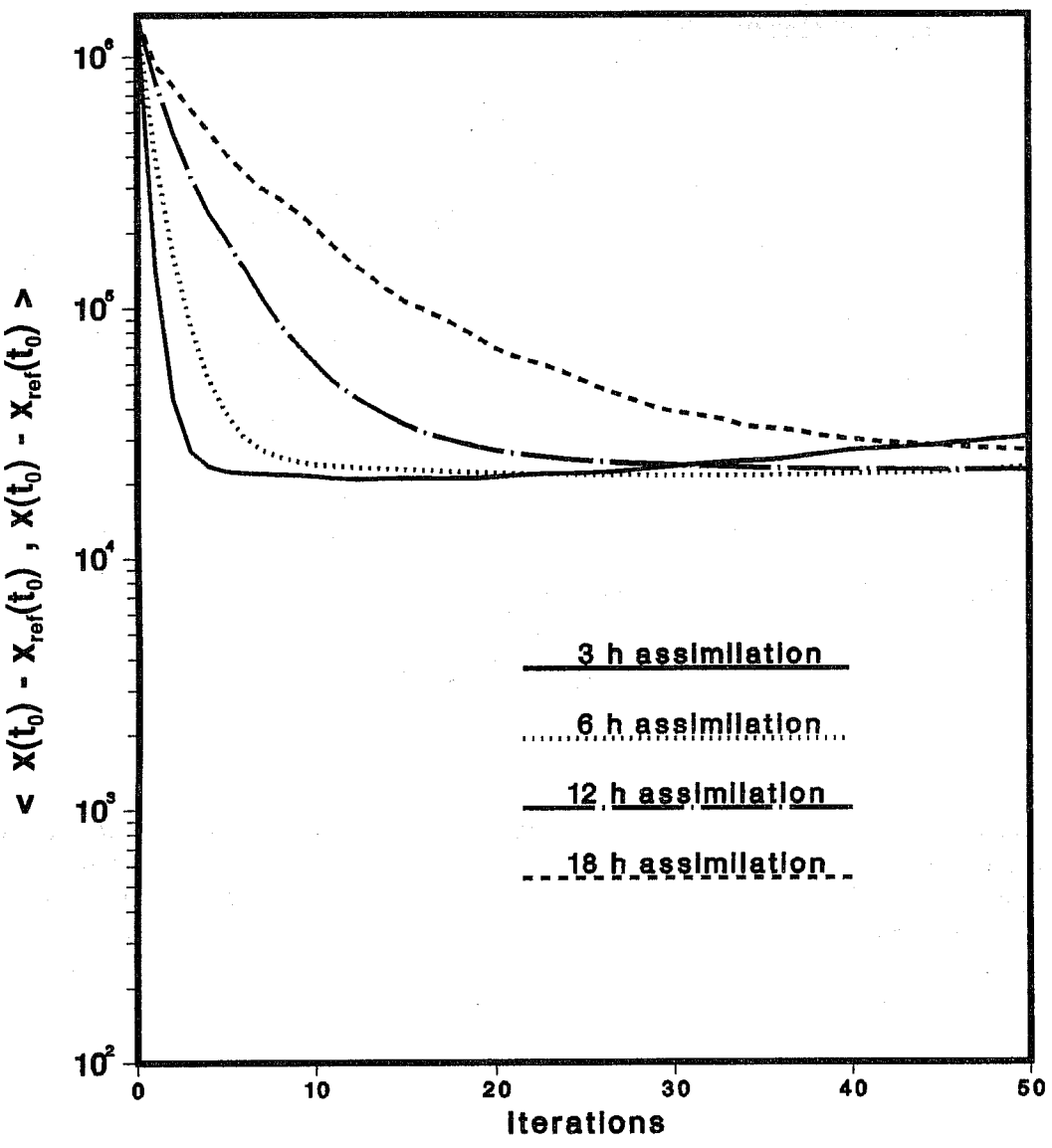


Fig. 12 Same as Fig. 3 but for different time periods of assimilation.

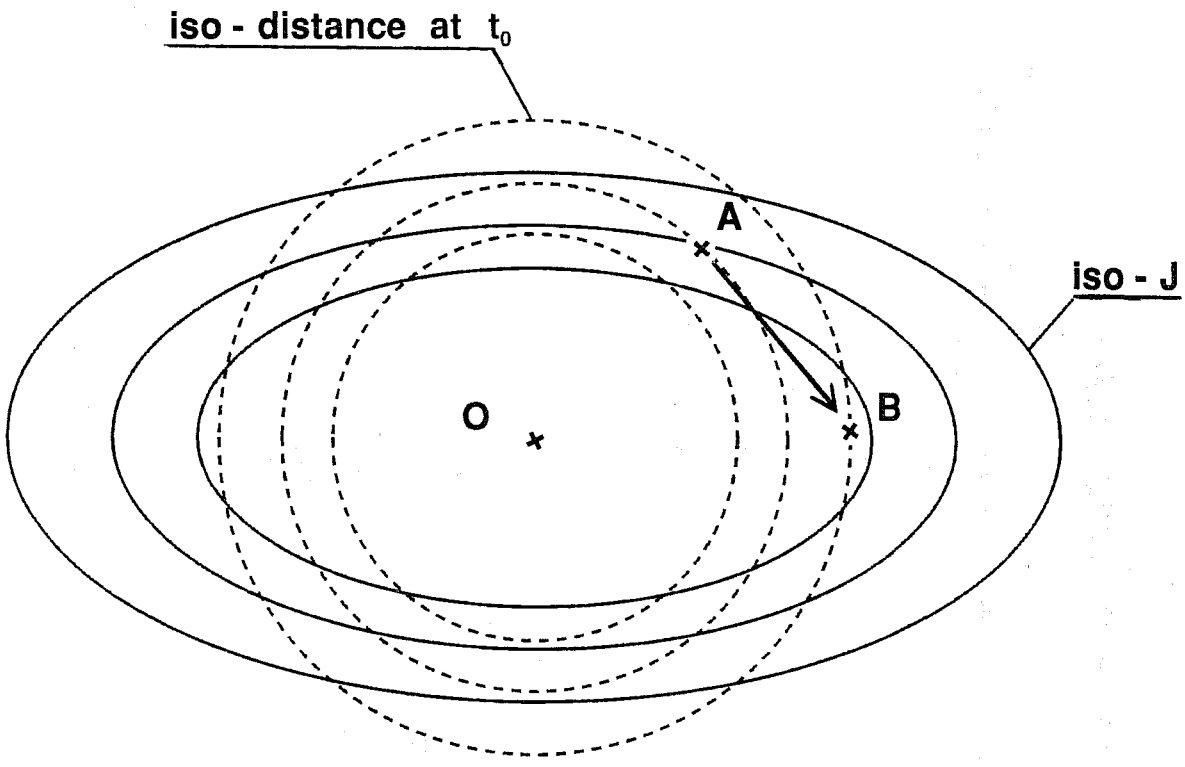


Fig. 13 Schematic 2-D representation of the cost-function and of the distance between the model solution and the reference state at time t_o. The inner product (for which the iso-distances at time t_o are spherical) has no reason to be the Hessian of the cost-function since the dynamics have no reason to conserve this particular inner product. The iso-J are then elliptic if we assume the validity of the tangent-linear approximation and can even be more complex. It is then possible to minimise J e.g. from A to B whereas the corresponding distance in phase space at time t_o increases.

We perform exactly the same experiment as above for a 6 h assimilation period but switching the horizontal diffusion on in the model. Fig. 14 shows the variations of the distance between the analysed field and the reference field with the number of iterations of the minimisation. The experiment with horizontal diffusion is represented in full line and the basic experiment in dashed line (already shown in Fig. 8 and Fig. 12). The convergence is much slower and the accuracy of the retrieval of the reference field is strongly affected.

Due to horizontal diffusion all the trajectories are converging toward the same solution and there is a loss of information in the course of the integration of the model. The problem of inverting the model is then ill conditioned in the same way that the heat equation is ill posed while integrated backward in time. Fig. 15 shows that the efficiency of the inversion is more altered for the upper levels of the model and for the small scales, i.e. where the horizontal diffusion has the strongest effect.

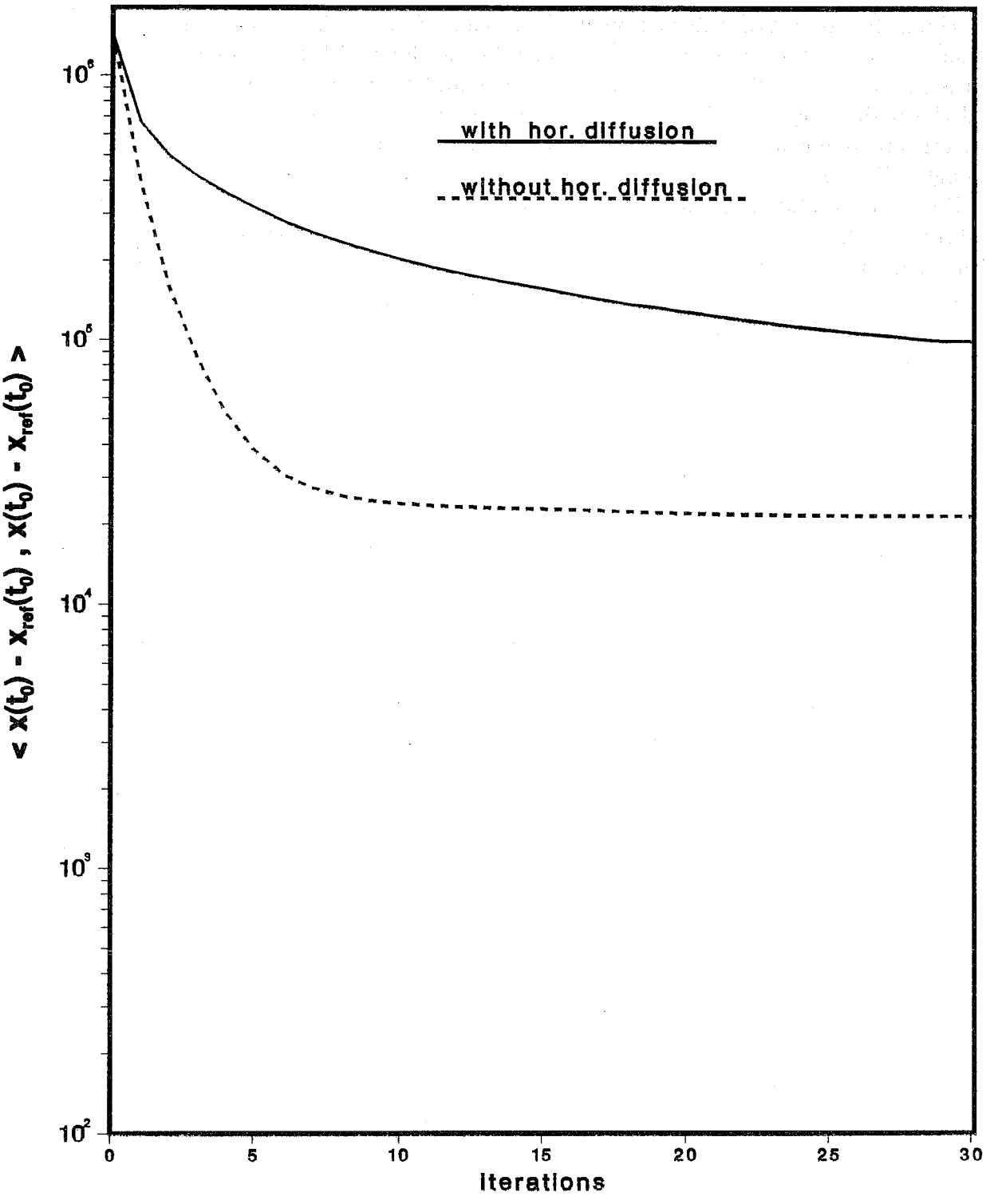


Fig. 14 Variation of the distance between the model state and the reference state with the number of iterations. Cases of assimilation with (solid curve) and without (dashed curve) horizontal diffusion.
(The dashed curve is the same as the solid curve in Fig. 8).

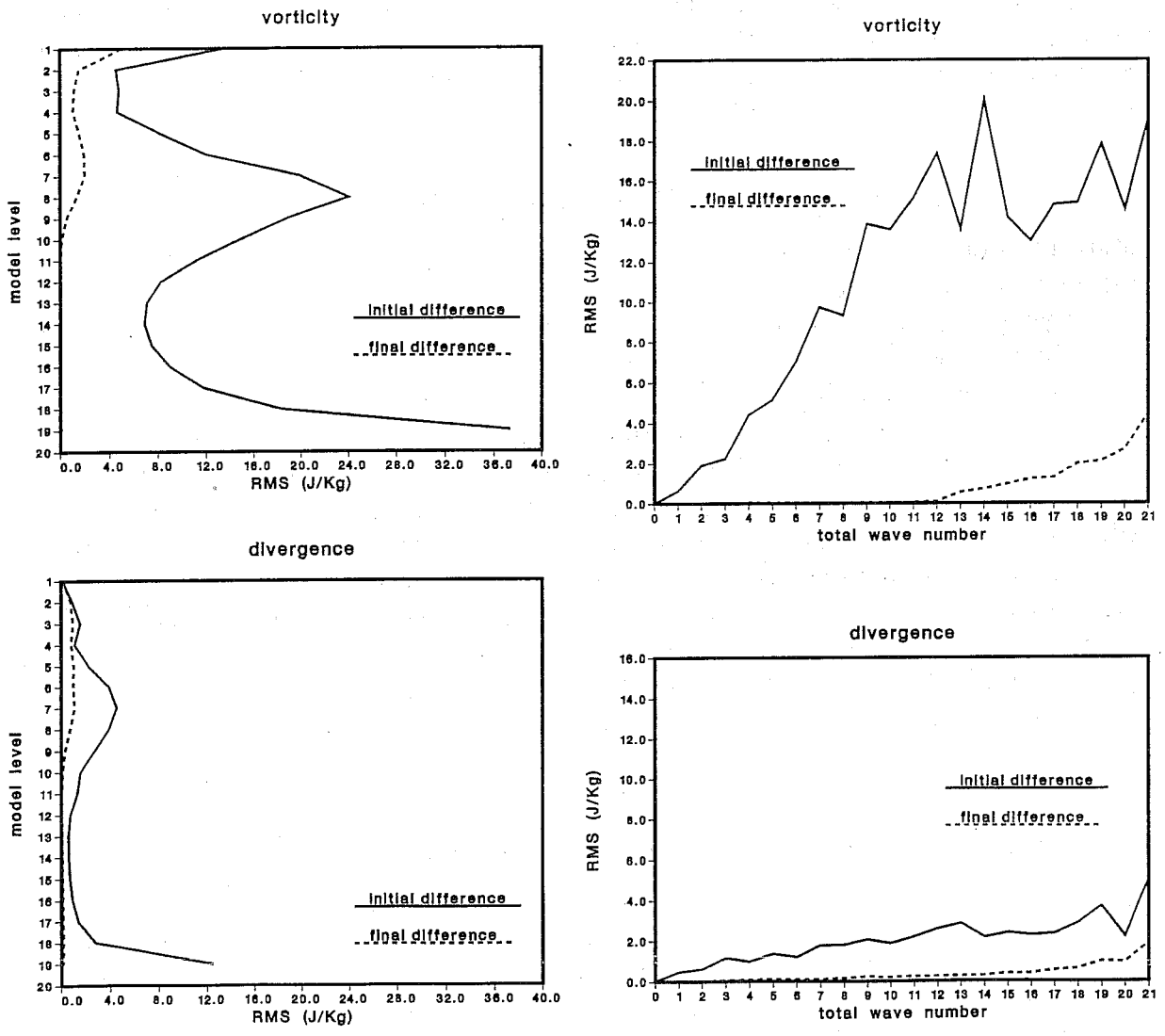


Fig. 15 Same as Fig. 5 and 6 but the model contains horizontal diffusion.

6 OBSERVATIONS DISTRIBUTED IN TIME.

The second set of experiments consists in studying the impact of having a set of observations distributed in time on the quality of the assimilation. Indeed, this context allows the evaluation of the ability of a full 4-D variational assimilation to use the information contained in the dynamics. The experiments presented below follow the same framework as those described previously. The reference state and the initial state of the minimisation are the same as in the first part of section 4.2. We perform an assimilation over a time period $[t_0, t_n]$ with the cost function given by (22).

6.1 Impact of the number of observations.

Starting from the experiment presented in Fig. 7, we have studied the impact of three different sets of observations :

- observations at t_3 and t_6,
- observations at t_0, t_3 and t_6,
- observations every hour.

As expected, the more we add observations, the smaller the saturation level in the reconstruction of the reference field is. In particular having observations at time t_0 before NNMI has a dramatic impact on the decrease of the cost function since information on the whole field is effectively available and not only the Rossby part. The difference between the experiment with observations available at t_0, t_3 and t_6 and the one with observations every hour is small. In both cases, the minimum of the cost function is reached in about 20 descent steps. Moreover, the distance at t_0 between the reference state and the result of the minimisation decreased by almost 4 orders of magnitude compared to 2 in the inversion experiments. The RMS of the differences is small for all fields (e.g. for the 500 hPa vorticity field it is equal to $10^{-8}s^{-1}$).

6.2 Information of the scales observed.

The scientific interest of the previous experiments is limited since all the fields are supposed to be observed everywhere and at different times. We shall study here the ability of 4-D variational assimilation to use the information contained in the dynamics together with observations of the fields. In this set of experiments, we assume that we observe only a limited part of the spectrum of the fields.

In the first experiment, the small scale features between total wave numbers 11 and 21 are observed every hour and we shall concentrate on the ability of the assimilation to recover the large scales. In Fig. 16 we present the differences by wave number at time t_0 between the model solution and the reference state for the vorticity and the divergence fields. The first remark is that, as could be expected the small scales are completely reconstructed by the assimilation. More interesting is that the large scale differences have decreased significantly for the vorticity field as well as for the divergence field.

In the second experiment, only the large scales are observed and in Fig. 17 we see that if satisfactory recovery of the large scale is achieved, the recovery of the small scale

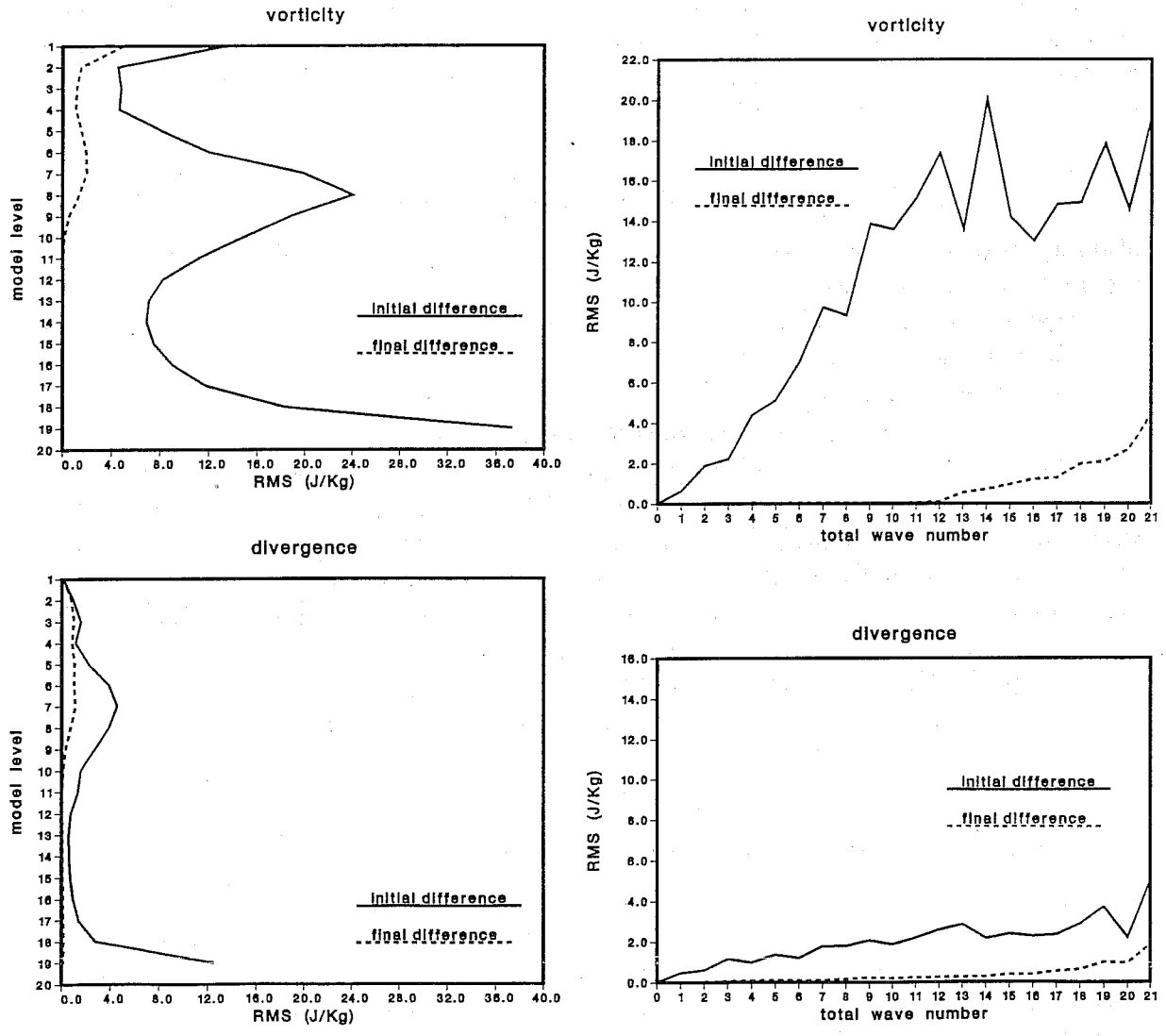


Fig. 15 Same as Fig. 5 and 6 but the model contains horizontal diffusion.

6 OBSERVATIONS DISTRIBUTED IN TIME.

The second set of experiments consists in studying the impact of having a set of observations distributed in time on the quality of the assimilation. Indeed, this context allows the evaluation of the ability of a full 4-D variational assimilation to use the information contained in the dynamics. The experiments presented below follow the same framework as those described previously. The reference state and the initial state of the minimisation are the same as in the first part of section 4.2. We perform an assimilation over a time period $[t_0, t_n]$ with the cost function given by (22).

6.1 Impact of the number of observations.

Starting from the experiment presented in Fig. 7, we have studied the impact of three different sets of observations :

- observations at t_3 and t_6,
- observations at t_0, t_3 and t_6,
- observations every hour.

As expected, the more we add observations, the smaller the saturation level in the reconstruction of the reference field is. In particular having observations at time t_0 before NNMI has a dramatic impact on the decrease of the cost function since information on the whole field is effectively available and not only the Rossby part. The difference between the experiment with observations available at t_0, t_3 and t_6 and the one with observations every hour is small. In both cases, the minimum of the cost function is reached in about 20 descent steps. Moreover, the distance at t_0 between the reference state and the result of the minimisation decreased by almost 4 orders of magnitude compared to 2 in the inversion experiments. The RMS of the differences is small for all fields (e.g. for the 500 hPa vorticity field it is equal to $10^{-8}s^{-1}$).

6.2 Information of the scales observed.

The scientific interest of the previous experiments is limited since all the fields are supposed to be observed everywhere and at different times. We shall study here the ability of 4-D variational assimilation to use the information contained in the dynamics together with observations of the fields. In this set of experiments, we assume that we observe only a limited part of the spectrum of the fields.

In the first experiment, the small scale features between total wave numbers 11 and 21 are observed every hour and we shall concentrate on the ability of the assimilation to recover the large scales. In Fig. 16 we present the differences by wave number at time t_0 between the model solution and the reference state for the vorticity and the divergence fields. The first remark is that, as could be expected the small scales are completely reconstructed by the assimilation. More interesting is that the large scale differences have decreased significantly for the vorticity field as well as for the divergence field.

In the second experiment, only the large scales are observed and in Fig. 17 we see that if satisfactory recovery of the large scale is achieved, the recovery of the small scale

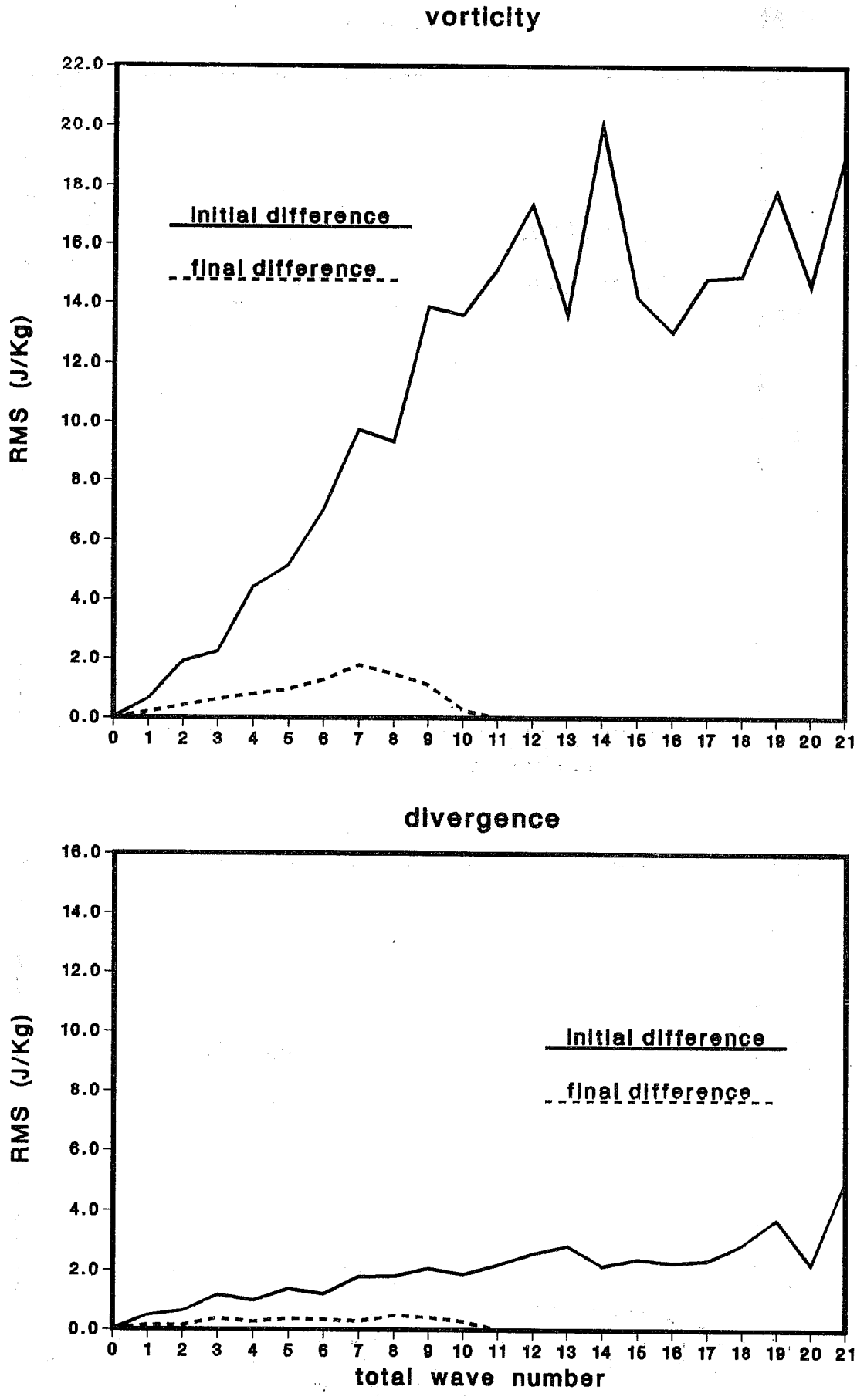


Fig. 16 Same as 6 but only the small scales are observed every hour.

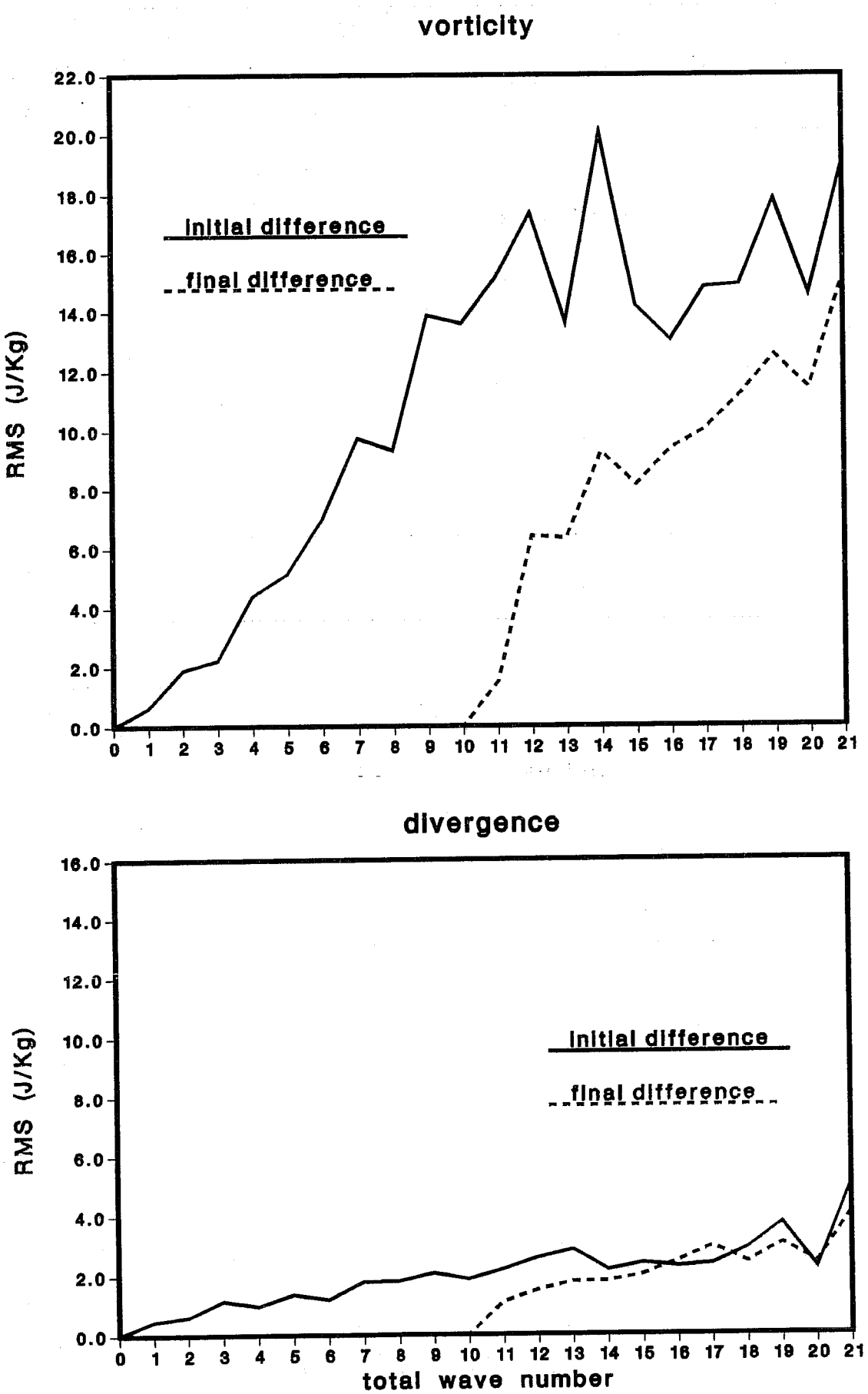


Fig. 17 Same as 16 but only the large scales are observed every hour.

features is weak, specially for the divergence field.

These results show that the cost function measuring the distance of the small scales is sensitive both to the small scales and to the large scales whereas in the second experiment the cost function measuring the distance of the large scales is sensitive to the large scales but far less to the small scales. This is consistent with the fact that the small scale structures behave over 6 hours to first approximation as passive tracers with little feedback on the large scale structures. The characteristic time for the small scale information to contaminate the large scales is larger than this period of 6 hours, even if the fact that it has been possible to improve to some extent on the vorticity field shows that there are some interactions. It is also worth mentioning that the observation of the small scale structures correspond in the physical space to more observations than the observation of the large scales.

These results validate the efficiency of the 4-D variational approach to use the information contained in the dynamics together with observations to infer components of the flow which are not directly observed. We have shown here an example of complementarity between small scale observations and dynamics to retrieve the large scale of the flow.

6.3 Information of the fields observed.

In connection with the previous experiments, we have performed several experiments in which only some of the meteorological fields were supposed to be observed.

For instance, we have performed an assimilation over 6 hours, the observations consisting in temperature and surface pressure fields every hour. We represent in Fig. 18 and Fig. 19 the differences between the reference state and the model solution at 500 hPa before and after minimisation and for respectively the vorticity field and the divergence field. It is worth noticing that, concerning the vorticity field, the differences are dramatically reduced in the mid-latitudes, whereas the minimisation is inefficient in the tropical belt. Two factors contributes to this result.

Firstly for the first five vertical modes, there is a one-to-one relationship between the mass and the vorticity field because of NNMI. Secondly, as already shown by Talagrand (1981) with a simpler data assimilation scheme and a simpler model, the knowledge of the mass field evolution is sufficient to recover to a certain extent the vorticity field in the mid-latitudes whereas the interaction between the two fields is reduced due to the nullity of the Coriolis force. In addition, we have performed the same experiment but with no NNMI and we obtain a similar result, i.e. observation of the mass field in the mid-latitudes is sufficient to reconstruct the vorticity part of the wind field. This clearly shows that this information is contained in the dynamics and is consistently used in our 4-D variational scheme.

As far as the divergence is concerned, we can also notice an important reduction of the differences. 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.

Another scenario consists of observations only in the troposphere. The results show that for an assimilation of 6 hours, the differences are reduced in the stratosphere. However, the period of assimilation was too short for the vertical coupling to be efficient and the reduction remains small.

8 ACKNOWLEDGMENTS.

The authors would like to thank Olivier Talagrand who took an active part in the planning and in the scientific discussions about this work. It is also a pleasure to thank Jean Pailleux who contributed with helpful advices and stimulating discussions. Mats Hamrud coded most of the adjoint of the model.

9 APPENDIX.

We intend to estimate the order of magnitude of the two terms on the right hand side of (8) given the experimental fact described in the text that the magnitude of the forecast errors in operational practice is similar to the magnitude of the observation errors. As the dynamics of the atmosphere shows an amplification of the errors with a typical doubling time between 1 and 2 days, we model the atmosphere with a single variable x governed by the linear dynamics :

$$x^{n+1} = \lambda x^n + \epsilon \tag{25}$$

where ϵ is a random variable which describes the model generated errors. Considering that x^{n+1} is six hour later than x^n, λ verifies the inequality : $\sqrt[8]{2} \leq \lambda \leq \sqrt[4]{2}$. (8) becomes :

$$P_g^{n+1} = \lambda^2 P_a^n + Q \tag{26}$$

Assuming we have one observation of x every six hours, which is reasonable over the continents comparing the scales resolved by the model and by the observational network, (12) becomes :

$$(P_a^n)^{-1} = (P_g^n)^{-1} + O^{-1} \tag{27}$$

Combining (26) and (27), one gets the following recurrence formulae :

$$P_g^{n+1} = \lambda^2 \left\{ (P_g^n)^{-1} + O^{-1} \right\}^{-1} + Q \tag{28}$$

and the limits P_g^∞ of P_g^n satisfies :

$$P_g^\infty = \lambda^2 \left\{ (P_g^\infty)^{-1} + O^{-1} \right\}^{-1} + Q \tag{29}$$

Using the experimental fact $P_g^\infty = O$, one gets the ratio between the two terms of (8) :

$$\frac{\lambda^2 (P_g^\infty)}{2Q} = \frac{\lambda^2}{2 - \lambda^2} \tag{30}$$

for the values of λ given above this ratio lies between 1.46 and 2.41. The amplification of the initial errors explains then something in between 60% in the worst case and 71% in the best case of the second hand of (8). The model generated errors explains only between 29% and 40% of (8). In a situation where the forecast errors were 50% greater than the observation errors, the amplification of errors would explain only between 47% and 57% of (8).

10 REFERENCES.

Buckley, A. and A. Lenir, 1983. QN-like Variable Storage Conjugate Gradients. *Mathematical Programming*, **27**, 155-175.

Caines P.E., 1988. Linear stochastic systems. *Wiley Series in Probability and Mathematical Statistics.*

Courtier, P., 1987. 'Application du contrôle optimal à la prévision numérique en météorologie.' Thèse de Doctorat de l'Université de Paris VI.

Courtier,P. and J.F.Geleyn, 1988. A global spectral model with variable resolution - application to the shallow water equations. *Q. J. R. Meteorol. Soc.*, **114**, 1321,1346.

Courtier, P. and O. Talagrand, 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation - Part II. Numerical results. *Q. J. R. Meteorol. Soc.*, **113**, 1329-1347.

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

Daley, R., 1978. Variational non-linear normal mode initialization. *Tellus*, **30**, 201-218.

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

Durand, Y., 1985. The use of satellite data in the French high resolution analysis. ECMWF workshop proceedings on high resolution analysis. 89-128.

Fillion, L. and C. Temperton, 1989. Variational implicit normal mode initialization. *Mon. Wea. Rev.*, **117**, 2219-2229.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube and E. Isaacson, 1981. Applications of Estimation Theory to Numerical Weather Prediction. In : *Dynamic meteorology. Data assimilation methods* (eds. L. Bengtsson, M. Ghil and E. Källén). Springer-Verlag, New-York, 139-224.

Ghil, M. and P. Malanotte-Rizzoli, 1990. Data assimilation in Meteorology and Oceanography. *To appear in Adv. Geophys.* .

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

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

Jarraud, M. and A.J. Simmons, 1983. The spectral technique. 1983 ECMWF seminar proceedings on numerical methods for weather prediction.

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

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.*, **82D**, 95-108.

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

Leith, C. E., 1980. Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.*, **37**, 954-964.

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

Machenhauer, B., 1977. On the dynamics of gravity oscillations in a shallow-water model with applications to normal mode initialization. *Beitr. Phys. Atmos.*, **50**, 253-271.

Mitchell, H. L., C. Charette, C. Chouinard and B. Brasnett, 1990. Revised interpolation statistics for the Canadian data assimilation procedure : Their derivation and application. *Mon. Wea. Rev.*, **118**, 1591-1614.

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

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

Talagrand, O., 1981. A study of the dynamics of four-dimensional data assimilation. *Tellus*, **33**, 43-60.

Talagrand, O. and P. Courtier, 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation - Part I. Theory. *Q. J. R. Meteorol. Soc.*, **113**, 1311-1328.

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

Tribbia, J. J., 1982. On variational normal mode initialization. *Mon. Wea. Rev.*, **110**, 455-470.