Tellus (1992), *44A*, 2–17

Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model

By PIERRE GAUTHIER, *Division de Recherche en Prévision Numérique, Atmospheric Environment Service, 2121 Transcanadian Highway, Dorval, P.Q., Canada H9P 1J3*

(Manuscript received 15 January 1991; in final form 7 June 1991)

ABSTRACT

Quadri-dimensional data assimilation aims at extracting all information from observations distributed over a finite time interval. In this paper, variational assimilation with the adjoint model technique is applied to the Lorenz model to illustrate how the performance of quadri-dimensional data assimilation can vary from one case to another. Observations are generated for two situations, one (the regular case) being more predictable than the other (the case with transition). An examination of the functional being minimized shows that although the regular case does not reveal any significant secondary minimum, there are in the case with transition for which the point of convergence was seen to be highly dependent on the first guess. It was also observed that to pick the first guess on the underlying attractor of this dynamical system does not insure convergence to the true minimum. In the adjoint model technique, the gradient of the functional is obtained through a time integration of the adjoint model using the difference between the solution of the direct model and the observations. It is shown how to relate the observational error covariance matrix to the gradient error covariance matrix. This method is applicable to any model once its adjoint is available and can be used to provide an estimate of the accuracy of the final analysis. Applying it to the Lorenz model, it is shown that due to the different local error growth rates, the same observational error can lead to very different accuracies for the gradient vector.

1. Introduction

The purpose of quadri-dimensional data assimilation is to extract information from observations contained in a finite time interval in order to produce an analysis to be used in a numerical forecast model. Integrating a model for given initial conditions Z_0, the solution is compared against the observations and a functional $J(Z_0)$ is defined to measure the misfit between this solution and the observations. In variational assimilation, the initial conditions that minimize this functional are sought. This approach was first tried by Thompson (1969) but the complexity of the problem made it intractable at the time. The minimization of $J(Z_0)$ requires the means to calculate the gradient $\nabla J(Z_0)$, a key ingredient in any minimization algorithm such as the conjugate gradient or the quasi-newton methods (Navon and Legler, 1987). A direct approach leads to a prohibitive cost and can only be used for simple models (Hoffman, 1986). What made variational assimilation to progress was the adjoint model technique developed by Le Dimet and Talagrand (1986) and studied also by Lewis and Derber (1985). They showed that the gradient can be obtained at a reasonable cost that is comparable to the estimation of the functional itself. The method was successfully applied to the analysis of real observations in spectral barotropic models on the sphere by Courtier and Talagrand (1987, 1990) and in a quasi-geostrophic baroclinic grid point model by Derber (1987). At the present time, there is ongoing work in forecast centers to perform variational assimilation for operational weather forecasting purposes.

In this paper, quadri-dimensional data assimilation will be cast in terms of a least square fit in which the functional

$$J(Z_0)=\frac{1}{(N+1)}\sum_{i=0}^{N}\langle Z(t_i)-\hat{Z}_i|Z(t_i)-\hat{Z}_i\rangle$$

has to be minimized. The following notations have been introduced: the inner product is $\langle f \mid g \rangle = f^{\mathrm{T}} g$ (the superscript $^{\mathrm{T}}$ stands for the matrix transpose), $Z(t)$ is the solution to the dynamical system

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = F(Z), \tag{1}$$

obtained with initial conditions $Z(0) = Z_0$ while $\hat{Z}_i$ are observed values of Z at time t_i. It has also been assumed that $Z \in \mathrm{R}^n$ and $F: \mathrm{R}^n \to \mathrm{R}^n$. An infinitesimal change δZ_0 to the initial conditions leads to a different trajectory in phase space and the evolution of the difference $\delta Z(t)$ between this solution and $Z(t)$ is described to first order by the *tangent linear model*

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta Z = DF(Z(t))\,\delta Z \equiv A(t)\,\delta Z, \tag{2}$$

with $DF(Z(t))$, the jacobian matrix of $F(Z)$. Insofar as δZ_0 is small enough, the trajectories remain close over the time interval $[t_0, t_N]$. In the adjoint model technique, an expression for the variation of $J(Z_0)$ is sought in the limit $\|\delta Z_0\| \to 0$ to obtain the gradient of the functional. In this limit,

$$\delta J(Z_0) = \langle \nabla J(Z_0) \mid \delta Z_0 \rangle$$

and this holds for any change in the initial conditions. Since

$$\delta J(Z_0) = \frac{2}{(N+1)} \sum_{i=0}^{N} \langle Z(t_i) - \hat{Z}_i \mid \delta Z(t_i) \rangle,$$

and the evolution of $\delta Z(t_i)$ being determined from (2), Le Dimet and Talagrand (1986) showed that

$$\nabla J(Z_0) = \frac{2}{(N+1)} \sum_{i=0}^{N} Q_i(Z(t_i) - \hat{Z}_i), \tag{3}$$

where Q_i stands for the backward integration from t_i back to t_0 of the adjoint model

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta^* Z = -A^{\mathrm{T}}(t)\,\delta^* Z. \tag{4}$$

For nonlinear systems, the divergence of trajectories in phase space is expected to vary from one region to another: this has an impact on the local predictability (Lorenz, 1982). This implies that if a model is to be integrated up to a given time, the accuracy needed on the initial conditions to obtain a good forecast should be expected to vary with the region of phase space. In meteorological terms, some situations being very sensitive to small changes in the atmospheric state, a detailed analysis is required while other cases can do with a less accurate one because the atmosphere can sometimes be very stable. The purpose of the present paper is to point out some difficulties that may arise in quadri-dimensional data assimilation. This will be done by using the Lorenz model (Lorenz, 1963) that has only three spectral components. It has been chosen because of its known sensitivity to small changes in the initial conditions: its properties have been investigated by many authors and its description is now part of textbooks on dynamical systems (e.g., Guckenheimer and Holmes, 1983).

There are many questions about data assimilation that can be answered by using simple models. Miller and Ghil (1990) tested the extended Kalman filtering method on the same Lorenz model that will be used in this paper while Lorenc (1988) used a unidimensional shallow water model to compare results obtained from quadri-dimensional data assimilation against methods used in operational practice (i.e., optimal interpolation in an analysis-forecast cycle). Data assimilation with the adjoint model technique being based on an iterative minimization process, results obtained from minimization algorithms such as the conjugate gradient or quasi-newton methods will depend on the starting point of the search if multiple minima are present. Moreover, it is to be expected that the form of J should vary with the region of phase space making the performance of the method case dependent. Another issue that will be looked at in this paper is the uncertainty in the gradient vector due to observational error. Compared to the true gradient that would be obtained from perfect observations, the gradient vector computed with a given set of observations would be different and this difference depends on the local error growth rate since, as can be seen from (3), a backward time integration of the adjoint model is involved.

In Section 2, the Lorenz model is introduced and the numerical schemes employed to integrate both the direct and adjoint models are discussed in

the Appendix. Observations are then generated by the model itself and data assimilation with the adjoint method is done for different cases. A conjugate gradient algorithm and a quasi-newton one are used for the minimization to compare their relative performance. Different cases are examined and for each one, a representation of the functional is given. The impact of observational error on the accuracy of the computed gradient is studied in Section 3. In Section 4, a discussion of the implications of these results for real data assimilating systems is presented.

2. Data assimilation with the Lorenz model

Studying a simplified form of the system proposed by Saltzmann (1962), Lorenz (1963) described the dynamics of finite amplitude convection in terms of a set of three nonlinear spectral equations that are now referred to as the *Lorenz model*. These equations are

$$\frac{\mathrm{d}}{\mathrm{d}t} X_1 = \sigma(-X_1 + X_2), \tag{5a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} X_2 = -X_1 X_3 + rX_1 - X_2, \tag{5b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} X_3 = X_1 X_2 - bX_3, \tag{5c}$$

where σ, r and b are external parameters related to the Prandtl and Rayleigh numbers and a geometric aspect ratio. A linear stability analysis shows that different types of solutions can be obtained depending on the values of these parameters. When $\sigma = 10$ and $b = \frac{8}{3}$, steady convection becomes unstable when $r \cong 24.74$. Choosing the Rayleigh number to be slightly supercritical ($r = 28$), it can be shown that the system then has an underlying strange attractor and the solutions are seen to be aperiodic (or chaotic). The reader is referred to Lorenz's original paper and the book by Guckenheimer and Holmes for a thorough discussion of the properties of this system. It should be mentioned that a similar behavior has been observed in other contexts. In relation to baroclinic instability, Pedlosky and Frenzen (1980) showed that similar equations can describe the dynamics of finite amplitude baroclinic waves in a quasi-geostrophic two-layer f-plane model. In Lorenz (1980, 1990) and Legras and Ghil (1985), chaotic behavior is shown to be emerging in forced-dissipative barotropic and baroclinic flows.

This system being very sensitive to small changes in the initial conditions, a small error on these implies that it is possible to make an accurate forecast only up to a finite period of time. Predictability can be measured by the time it takes for the error to grow up to a prescribed level and it usually varies from case to case because the local error growth rate does (Lorenz, 1982; Benzi and Carnevale, 1989). The complex behavior of this simple dynamical system makes it an ideal tool to test ideas quickly for anyone with moderate computer resources. In identical twin experiments, observations are generated by the model itself and noise is added (or not) to simulate observational error.

The impact of chaos on data assimilation will be studied with the parameter values $\sigma = 10$, $b = \frac{8}{3}$

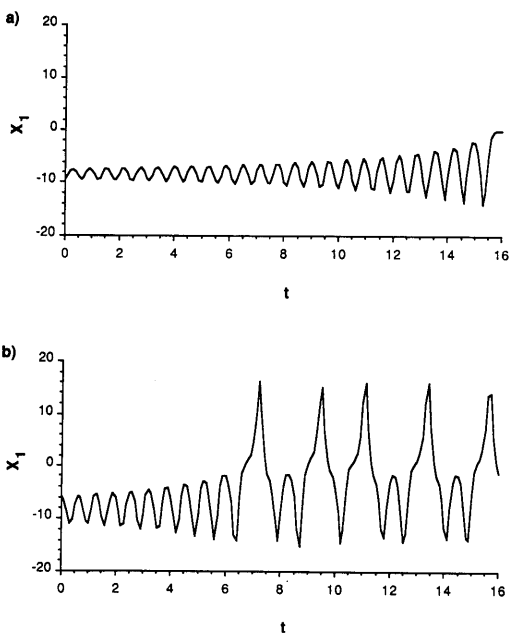

Fig. 1. Representation of the component $X_1(t)$ as a function of time: (a) Regular case generated from the initial conditions $Z_0 = (-9.42, -9.34, 28.3)$, (b) Case with transition generated from $Z_0 = (-5.92, -5.90, 24.0)$. Data assimilation is performed over the time interval $0 < t < 8$ and a forecast is made from $t = 8$ up to $t = 16$.

and $r = 28$ in the Lorenz model. Identical twin experiments will be performed and to insure that observations are on the attractor, the model is run for some time (5 time units) before picking up observations. If $Z \equiv (X_1, X_2, X_3)^T$, the adjoint model is given by (4) with

$$A(t) = \begin{pmatrix} -\sigma & \sigma & 0 \\ r - X_3(t) & -1 & X_1(t) \\ X_2(t) & X_1(t) & -b \end{pmatrix}, \tag{6}$$

and the gradient is obtained from (3) with Q_i replaced by its discretized equivalent (see the Appendix). Two cases were chosen: one for which the solution is regular over the finite time interval $t \in (0, 16)$ and another that experiences a transition of regime. They were generated from initial conditions $(-9.42, -9.34, 28.31)$ and $(-5.92, -5.90, 24.0)$ respectively and are shown on Fig. 1. Data assimilation with the adjoint model technique was performed over the time interval $[0, 8]$ and the resulting analysis is used to make a forecast from $t = 8$ up to $t = 16$. For these two cases, observations are taken as the exact solutions at $t = 0$, 2, 4, 6 and 8. The minimization of $J(Z_0)$ has been performed with both a quasi-newton algorithm (subroutine E04KBF from the NAG library) and a conjugate gradient (subroutine ZXCGR from the IMSL library).

For the regular case, several experiments were conducted with different starting points Z_0 for the minimization and for $Z_0 = (-8.0, -8.0, 29.0)$, Fig. 2 represents the value of J versus the number of function evaluations as the minimization was carried on with a conjugate gradient or a quasi-newton. In the early stages, both algorithms have a comparable performance. However, in the final stages, the iterations have brought the analyzed point in the neighborhood of the true minimum. Therefore, the quadratic approximation of the functional becomes more reasonable and the quasi-newton method is able to estimate the hessian matrix correctly from the values of the gradient calculated at different points. Having a correct estimation of the quadratic part, the quasi-newton method gains some speed over the conjugate gradient method that does not use this information as efficiently. The convergence was achieved in both cases and the final value of the functional was virtually zero ($\sim 10^{-6}$) so that the correct initial conditions were recovered perfectly (to within the truncation order of the numerical quadrature). It follows then that the forecast made from this analysis was also nearly perfect over the forecast time interval. In practice, however, it often occurs that the process must be stopped when a maximum number of iterations has been reached. To see the impact of an imperfect analysis on the forecast, an analyzed state for which $J(Z_0) = 0.037$ was picked at some point during the minimization process. Fig. 3 shows the resulting analysis and

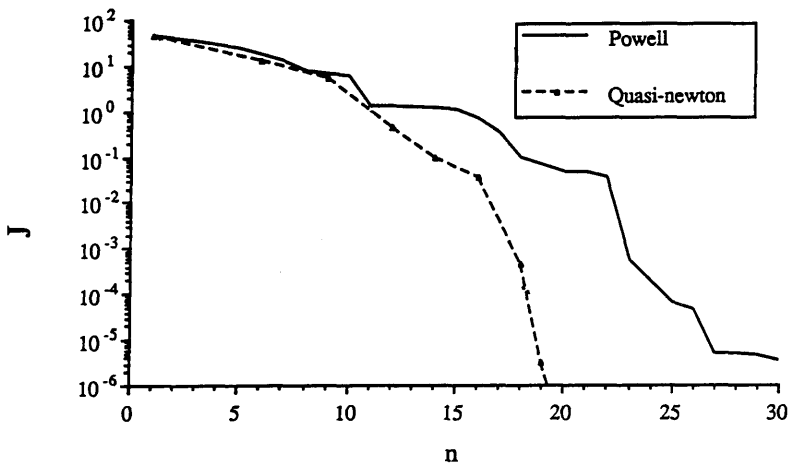

Fig. 2. Value of the cost function $J(Z_0)$ as a function of iterations for a Powell conjugate gradient algorithm (solid line) and a quasi-newton algorithm (dashed line).

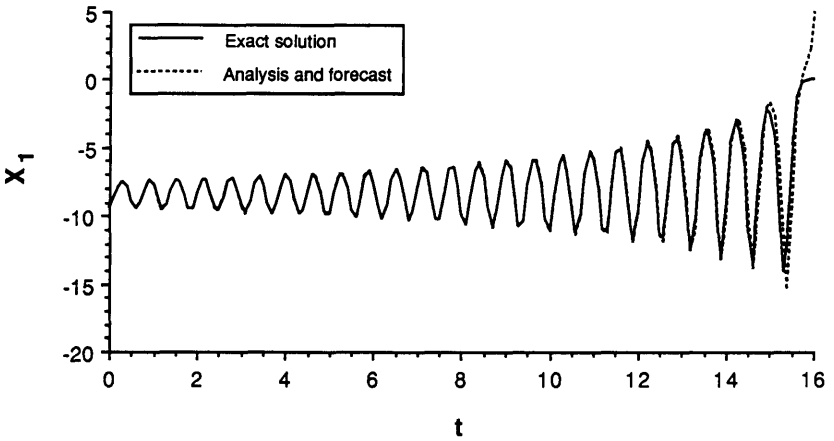

Fig. 3. Component $X_1(t)$ of the analysis and the resulting forecast for the regular case. The imperfect analysis is taken from an intermediate step of the minimization process where $J(Z_0) = 0.037$.

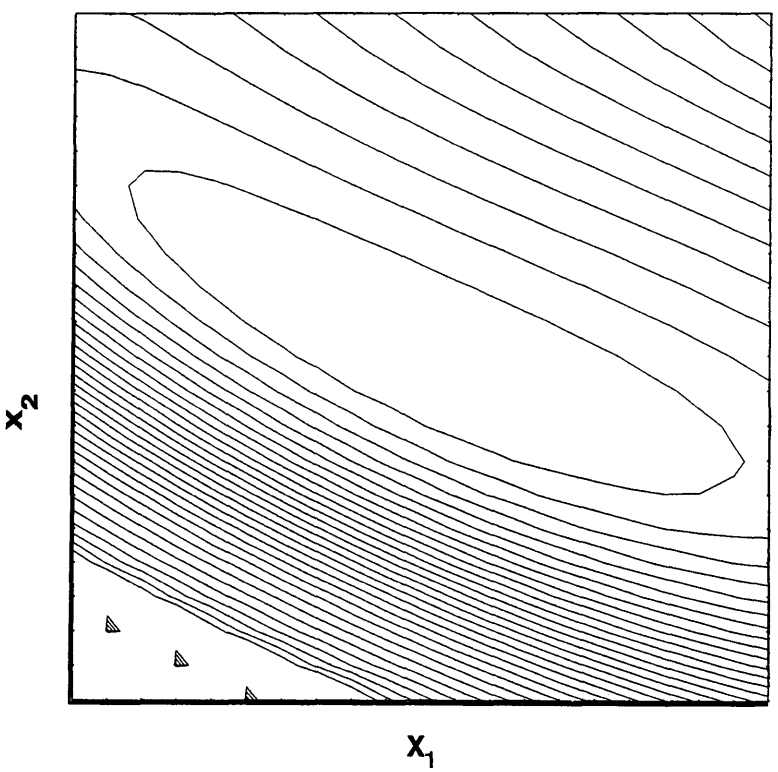

Fig. 4. Representation of the functional $J(Z_0)$ for the regular case: $Z_0 = (X_1, X_2, X_3)$ was varied such that:

$X_1^* - 2 < X_1 < X_1^* + 2,$

$X_2^* - 2 < X_2 < X_2^* + 2,$

with $X_3 = X_3^*$ and $(X_1^*, X_2^*, X_3^*) = (-9.42, -9.34, 28.3)$. Contour intervals are evenly spaced for values of $J(Z_0)$ between 0 and 50.

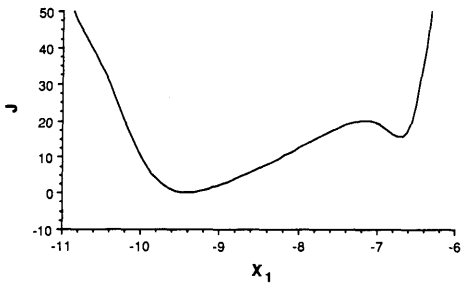

Fig. 5. Values of $J(Z_0)$ versus X_1 along a straight line going through the points

$P_{\pm} = (X_1^* \pm 2, X_2^* \pm 2, X_3^*)$.

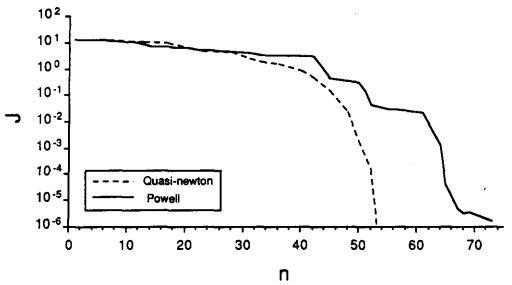

Fig. 6. Same as Fig. 2 but for the case with transition.

forecast. Even though the error in the analysis is not discernable, it results nevertheless in a forecast error near the end of the forecast period.

Having a nonlinear model, the question of the existence of multiple minima arises. To give a picture of $J(Z_0)$, the functional was evaluated by fixing $X_3(0)$ to its true value while $X_1(0)$ and $X_2(0)$ varied. From the contour levels of $J(Z_0)$ plotted on Fig. 4, it is apparent that the functional is fairly quadratic. If one is to use a reasonable first guess, the minimization converges to the true minimum. On the other hand, if the first-guess is far away from the minimum, the method may converge to a secondary one: this has happened while experimenting with different starting points. This may not be apparent on Fig. 4 but an indication of such a secondary minimum can be seen on Fig. 5 representing values of $J(Z_0)$ along a line going through the points

$P_{\pm} = (X_1^* \pm 2, X_2^* \pm 2, X_3^*)$

which corresponds to a cross-section along one diagonal of the region shown on Fig. 4. Since the phase space has more than one dimension, this may not be a minimum at all. But even if it were, this minimum would not be very significant since the point $X_1 = -7$ implies a relative error level of over 25% in the first-guess and one could argue that in practice, the first guess would not be that far from reality. From this first experiment, one therefore concludes that assimilation with the ajoint model approach has succeeded in providing a dependable analysis that leads to an accurate forecast.

The case with transition however is more of a test for quadri-dimensional data assimilation and

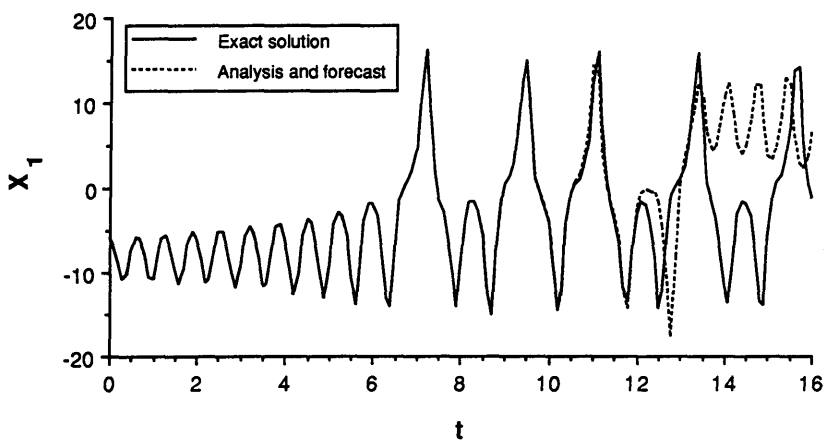

Fig. 7. Same as Fig. 3 but for the case with transition. The imperfect analysis is taken from an intermediate step where $J(Z_0) = 0.023$.

can be related to atmospheric events where the flow configuration experiences a sudden change. One could think of the formation and dissolution of a blocking regime or the development of a synoptic system through baroclinic instability. As was pointed out earlier, in both cases the dynamics can be linked to a chaotic regime similar to what the Lorenz model exhibits (see Pedlosky and Frenzen (1982) for baroclinic instability and Legras and Ghil (1985) for blocking). Using the point (−3.99, −7.50, 24.0) as a first guess, Fig. 6 shows the results of data assimilation for the case with transition generated with initial conditions (−5.92, −5.90, 24.0). Convergence to the right solution was obtained even though at a much slower rate than in the previous case. Again the quasi-newton method is improving the convergence in the latter stages while the conjugate gradient method has a comparable performance at the beginning. At the end of the assimilation, the value of the functional is nearly zero and the true solution was accurately recovered. The impact of using a good but not perfect analysis was looked at by choosing an analyzed state at some intermediate stage of the minimization process for which $J(Z_0) = 0.023$ which is comparable to what was considered for the regular case to obtain Fig. 3. The analysis and the resulting forecast are represented on Fig. 7: they indicate that this situation is less predictable than the previous one since for a similar error on the initial conditions, the forecast ceases to be accurate sooner than in the regular case.

The most important result for the case with transition is that convergence to secondary minima was obtained several times while experimenting with different starting points for

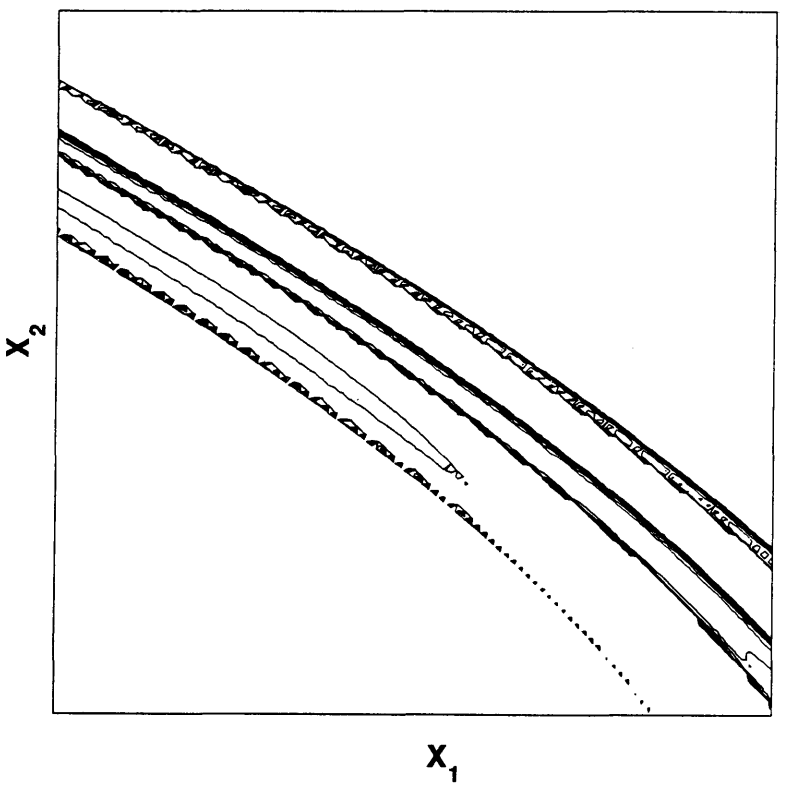

Fig. 8. Same as Fig. 4 but for the case with transition with

$(X_1^*, X_2^*, X_3^*) = (-5.92, -5.90, 24.0)$.

Contour intervals are evenly spaced for values of $J(Z_0)$ between 0 and 50. Values of $J(Z_0)$ above 50 are not shown.

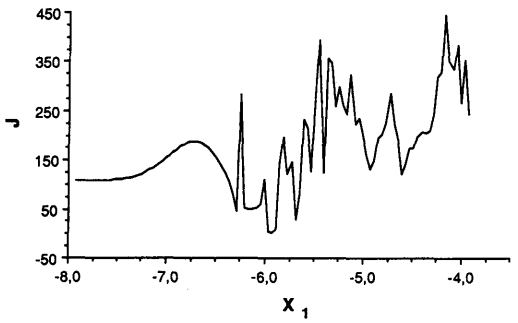

Fig. 9. Same as Fig. 5 but for the case with transition with

$(X_1^*, X_2^*, X_3^*) = (-5.92, -5.90, 24.0)$.

the minimization. To shed some light on this, the functional was again plotted by holding $X_3(0)$ fixed to its true value while $X_1(0)$ and $X_2(0)$ were varied; Fig. 8 shows the resulting contour levels of $J(Z_0)$ with a contour interval identical to the one used in Fig. 4. Since values above 50 were discarded on this figure, a more complete picture is obtained by evaluating the functional on a cross-section of the region shown on Fig. 8 along the diagonal going through the points $P_{\pm}$ defined earlier (Fig. 9). Fig. 8 and 9 show that any minimization algorithm is bound to have problems because of the very narrow region around the absolute minimum where the functional can be well approximated by a quadratic function. Therefore, to obtain convergence a first guess that stands very close to the true solution should be used. The contour levels of Fig. 8 show that low values of the functional can be obtained at points that are not close to the true minimum. Measuring the error associated with a solution by $\|Z(t) - Z_t(t)\|$ with Z_t the true solution, Fig. 10 shows that the solution that used point $B = (-5.80, -5.80, 24.0)$ as initial conditions leads to a more important error at the end of the time interval than a solution that used point $A = (-3.99, -7.50, 24.0)$ as initial conditions: the relative position of these two points is shown on Fig. 11. The result is that the functional has a lower value at point A than point B even if this one is closer to the true solution. Moreover, using B as a first guess leads to convergence to a secondary minimum (its projection on the plane defined on Fig. 11 has been indicated as point C on this figure) while point A converged to the right one. In fact, convergence to a secondary minimum never occurred in several experiments that used first guesses having low values for the functional.

Since convergence to secondary minima is a real concern when performing data assimilation through minimization algorithms, it is worth looking at what makes the points having low values of $J(Z_0)$ immune to this problem: understanding why would give us an edge on picking up a good first guess. The first argument that comes to mind is that the data being located on the attractor of this dynamical system, the problem should

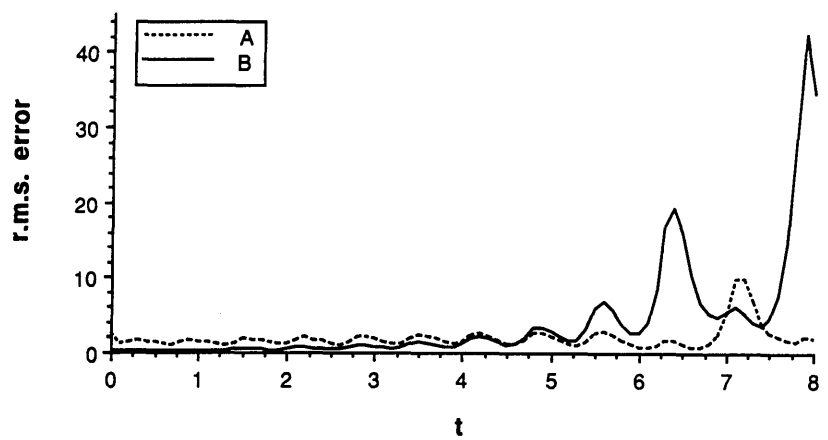

Fig. 10. Root-mean-square error for two solutions generated from different initial conditions in the case with transition: A corresponds to the initial conditions $(-3.99, -7.50, 24.0)$ and B to $(-5.80, -5.80, 24.0)$.

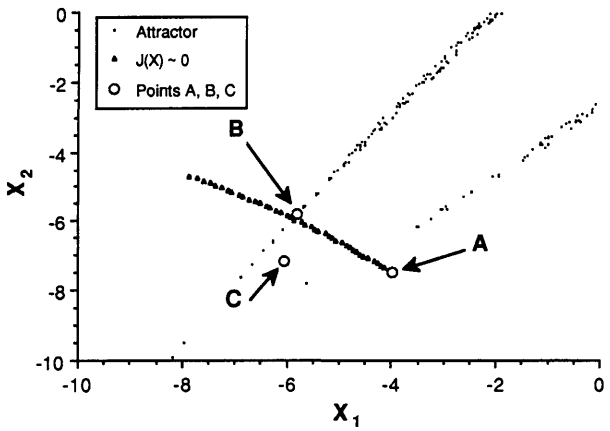

Fig. 11. Intersection of the Lorenz attractor with the plane used in Fig. 8. The contour level of low values of $J(Z_0)$ has been represented with larger dots. Points A and B discussed in Fig. 10 have also been indicated while point C corresponds to the projection of the secondary minimum to which converged a minimization using B as a first guess.

somehow be restricted to points located on it. While performing data assimilation with a barotropic shallow water model, Courtier and Talagrand (1990) pointed out that the constraint of being on the slow manifold had to be imposed to obtain a good analysis: in this context, the balance constraint characterizes the underlying attractor of the dynamics. Here, it is not possible to impose that initial conditions be on the Lorenz attractor or even approximately on it. However, it is possible to represent this attractor by plotting its intersection with a plane: this can be done by running the model for a long period of time (in climate mode) and marking the points each time the trajectory intersects a given plane. Employing the same plane used in Fig. 8, the intersection of the attractor was plotted on Fig. 11 and the contour level of low values of the functional was also indicated on it for reference because it is representative of those points that would converge to the true minimum. This contour level is nearly perpendicular to the attractor therefore implying that to choose a first guess on the attractor may not be worth the effort. This shows that the set of points corresponding to first guesses that lead to convergence to the absolute minimum can be different from the attractor associated with the dynamics.

As was pointed out to the author by Olivier Talagrand, this feature can be understood by the fact that in a chaotic dissipative system, the orbits will simultaneously converge toward the attractor while diverging on the attractor. Therefore, a perturbation of given amplitude along the attractor will lead to rapid divergence while a perturbation of similar amplitude but taken in a direction perpendicular to it will be damped as the perturbed solution relaxes back to the attractor. The fact remains that if it is known that the dynamics possesses an attractor, the final analysis should be on it or at least lead to an evolution that is close to it: this is exactly what happens here since the true solution lies at the intersection point between the contour level and the attractor.

3. Impact of observational error on the accuracy of the calculated gradient

In the following discussion, it will be assumed that there is no model error but observations $\hat{Z}_i$ are such that

$$\hat{Z}_i = Z_i^t + b_i,$$

where Z_i^t stands for the true value of Z at time t_i while b_i represents an unbiased observational error that is representative of both the measurement and representativeness inaccuracies. In the preceding section, observed values were assumed to be exact so that the functional

$$J_t(Z_0) = \frac{1}{(N+1)} \sum_{i=0}^{N} \langle Z(t_i) - Z_i^t | Z(t_i) - Z_i^t \rangle$$

vanishes when Z_0 corresponds to the correct analyzed state at $t=0$ (noted by Z_0^*).

By using imperfect data, the functional

$$J(Z_0)=\frac{1}{(N+1)}\sum_{i=0}^{N}\langle Z(t_i)-\hat{Z}_i|Z(t_i)-\hat{Z}_i\rangle$$

$$=J_t(Z_0)-\frac{2}{(N+1)}\sum_{i=0}^{N}\langle Z(t_i)-\hat{Z}_i|b_i\rangle$$

$$+\frac{1}{N+1}\sum_{i=0}^{N}\langle b_i|b_i\rangle \qquad (7)$$

and $J(Z_0^*)$ no longer necessarily vanishes.

By cumulating statistics, it is possible to have an idea of what the observational error is on average. If restricted to measurement error only, it is a reasonable assumption to state that b_i does not depend on the atmospheric state and also that the error involved between two different measurements are statistically independent. To assess the impact of this error on the functional, an ensemble average is made over many experiments using different sets of measurements for the same atmospheric case. Therefore for experiment "k", the observed values would be

$$\hat{Z}_{ik}=Z_i^t+b_{ik},$$

so that the true state is always the same. The ensemble average of (7) is then

$$\overline{J(Z_0)}=J_t(Z_0)+\frac{1}{N+1}\sum_{i=0}^{N}\overline{\langle b_i|b_i\rangle},$$

where the overbar stands for the ensemble average. At $Z_0=Z_0^*$, $\overline{J(Z_0^*)}$ does not vanish but is a measure of the total observational error variance.

The value of ∇J is at the heart of most minimization algorithms and it is then important to examine the impact of having imperfect data on its accuracy. From (3), one gets that

$$\nabla J(Z_0)=\nabla J_t(Z_0)+G'$$

where

$$G'=-\frac{2}{N+1}\sum_{i=0}^{N}Q_i b_i \qquad (8)$$

is the error made on ∇J due to observational error. Remembering that Q_i does not involve the observations, then

$$\overline{G'}=-\frac{2}{N+1}\sum_{i=0}^{N}\overline{Q_i b_i}$$

$$=-\frac{2}{N+1}\sum_{i=0}^{N}Q_i\overline{b_i}=0,$$

the error on the observations being assumed unbiased. Therefore, the ensemble average of ∇J corresponds to its exact value and the gradient is seen to be unbiased.

Because $G'=G'(Z_0)$, it also depends on the local error growth rate and as a consequence, it is to be expected that a given b_i could lead to a different error on ∇J. From (8), the gradient error covariance matrix is seen to be

$$\overline{G'G'^{\mathrm{T}}}=\frac{4}{(N+1)^2}\sum_{n=0}^{N}\sum_{m=0}^{N}Q_n\overline{b_n b_m^{\mathrm{T}}}Q_m^{\mathrm{T}}. \qquad (9)$$

A reasonable assumption about observational error is that it is uncorrelated in time, an assumption that is also made in the derivation of the Kalman filter equations (Ghil, 1989). Consequently,

$$\overline{b_n b_m^{\mathrm{T}}}=\delta_{nm}R_n$$

with δ_{nm}, the Kronecker delta function and R_n is the observational error covariance matrix at time t_n: (9) then boils down to

$$\overline{G'G'^{\mathrm{T}}}=\frac{4}{(N+1)^2}\sum_{n=0}^{N}Q_n R_n Q_n^{\mathrm{T}}. \qquad (10)$$

This relates the gradient error covariance matrix to R_n. The cost of performing this calculation corresponds to $2K$ integrations of the adjoint model with K being the number of model variables: this can be done by using the adjoint model alone and can be applied to any model insofar as its adjoint is available.

This relation is now applied to the Lorenz model for the two cases studied in Section 2. In all cases to be described below, R_n has been considered to be diagonal with its diagonal elements D being equal and the gradient error covariance matrix has been computed at the point $Z_0=Z_0^*$ where the true gradient vanishes. Data insertions were

regularly spaced and occurred at every 0.8 time units over the time interval $[0, 8]$ and $D = 0.01$. For the regular case, the result obtained was

$$\overline{G'G'^{\mathrm{T}}} = \begin{pmatrix} 0.006 & 0.010 & -0.004 \\ 0.010 & 0.023 & -0.005 \\ -0.004 & -0.005 & 0.005 \end{pmatrix}, \qquad (11)$$

while for the case with transition,

$$\overline{G'G'^{\mathrm{T}}} = \begin{pmatrix} 4.494 & 6.375 & -3.950 \\ 6.375 & 9.056 & -5.586 \\ -3.950 & -5.586 & 3.504 \end{pmatrix}. \qquad (12)$$

This reveals that the same observation error does not have the same impact on the accuracy of the gradient and this is a consequence of the different local error growth rates.

The error variance is obtained from the diagonal elements of this covariance error matrix and it then makes it possible to investigate for instance, the dependence of this error on the quality of the data and their geographical distribution. For the Lorenz model, the impact of different strategies on the accuracy of the gradient is looked at by either improving the quality of the observations or by increasing their number. Considering the case with transition, if the number of data insertions is increased from 11 up to 41 while keeping the accuracy fixed ($D = 0.01$), then

$$\overline{G'G'^{\mathrm{T}}} = \begin{pmatrix} 0.519 & 0.741 & -0.449 \\ 0.741 & 1.060 & -0.638 \\ -0.449 & -0.638 & 0.392 \end{pmatrix},$$

and this leads to a more accurate gradient. On the other hand, keeping the same number of data insertions fixed to 11 but increasing the quality of the observations by setting $D = 0.001$ gives the result that

$$\overline{G'G'^{\mathrm{T}}} = \begin{pmatrix} 0.449 & 0.638 & -0.395 \\ 0.637 & 0.906 & -0.559 \\ -0.395 & -0.559 & 0.350 \end{pmatrix}$$

and the gradient is also improved. However, one should note that in both cases, the gradient remains less accurate than it was in the regular case.

The trace of this covariance matrix gives a measure on the uncertainty there is on $\|\nabla J\|$ due to the presence of observational error. Since the criterion of convergence of minimization algorithms is usually based on the fact that $\|\nabla J\|$ is small enough, it is to be expected that in one realization, $\|\nabla J(Z_0^*)\|$ would not vanish and the analyzed state would correspond rather to a point Z_1^* where $\|\nabla J(Z_1^*)\|$ vanishes but the true gradient does not. To illustrate this argument, variational data assimilation with the adjoint model technique was done with a non-divergent barotropic model on the sphere, a model that was used in Talagrand and Courtier (1987) (TC hereafter): their paper gives a complete description of the adjoint equations altogether with details about the implementation of the technique in a pseudo-spectral model. Here, spherical harmonics Y_n^m normalized with respect to the inner product

$$\langle f \mid g \rangle = \frac{1}{2\pi} \iint fg \, \mathrm{d}\lambda \, \mathrm{d}\mu$$

are employed with a triangular truncation of order 21 and all harmonics (even and odd) are kept: the system has therefore 483 degrees of freedom. These equations were integrated with a leap-frog scheme and the first timestep is Eulerian: the adjoint of the discretized scheme (as described in TC) is used. Observations were kept at every timestep ($\Delta t = 30$ min) and assimilation is performed over a period of 12 h.

In the first experiment, observations were generated by integrating the model with the initial conditions

$$\zeta(\lambda, \mu, 0) = \alpha Y_1^0 + \beta Y_5^4,$$

for the vorticity field ζ with $\alpha = 1.187 \times 10^{-5}\ \mathrm{s}^{-1}$ and $\beta = 1 \times 10^{-5}\ \mathrm{s}^{-1}$: as usual, λ is the longitude and $\mu = \cos\theta$, θ being the colatitude. The solution corresponds to a Haurwitz wave moving with a phase speed of $9.55°\ \mathrm{day}^{-1}$ and this experiment is very similar to the one reported in TC except for the fact that only the odd harmonics were kept in their paper. The minimization of the functional was performed with the same conjugate gradient algorithm that was used in Section 2 and the first guess was chosen to be a random state generated

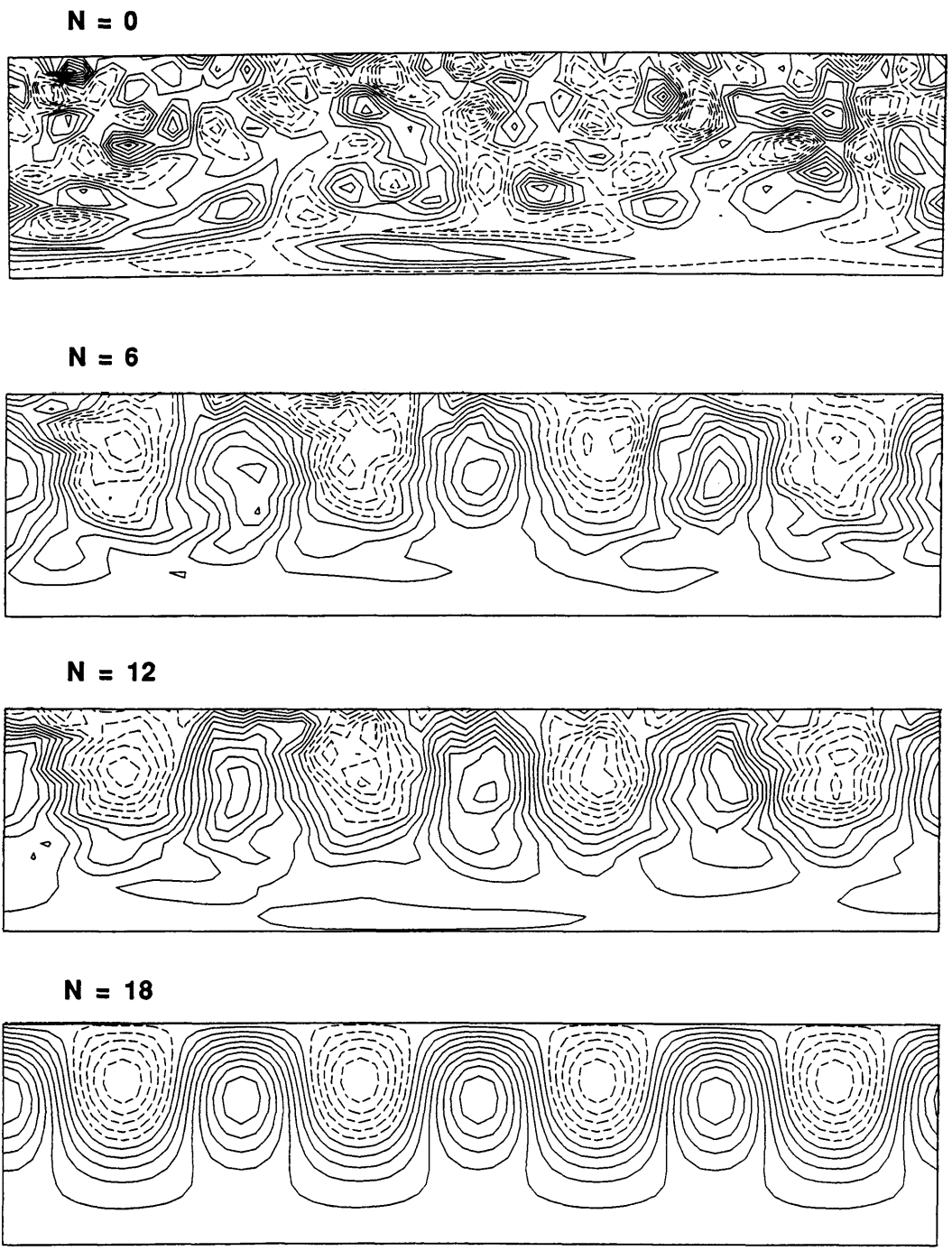

Fig. 12. Latitude-longitude representation of the vorticity field at $t = 0$ as a function of iterations for an assimilation that used perfect observations at every time step over a 12-h time interval. The domain is such that $0 < \lambda < 2\pi$ is the horizontal axis and $0 < \mu < 1$ is the vertical axis.

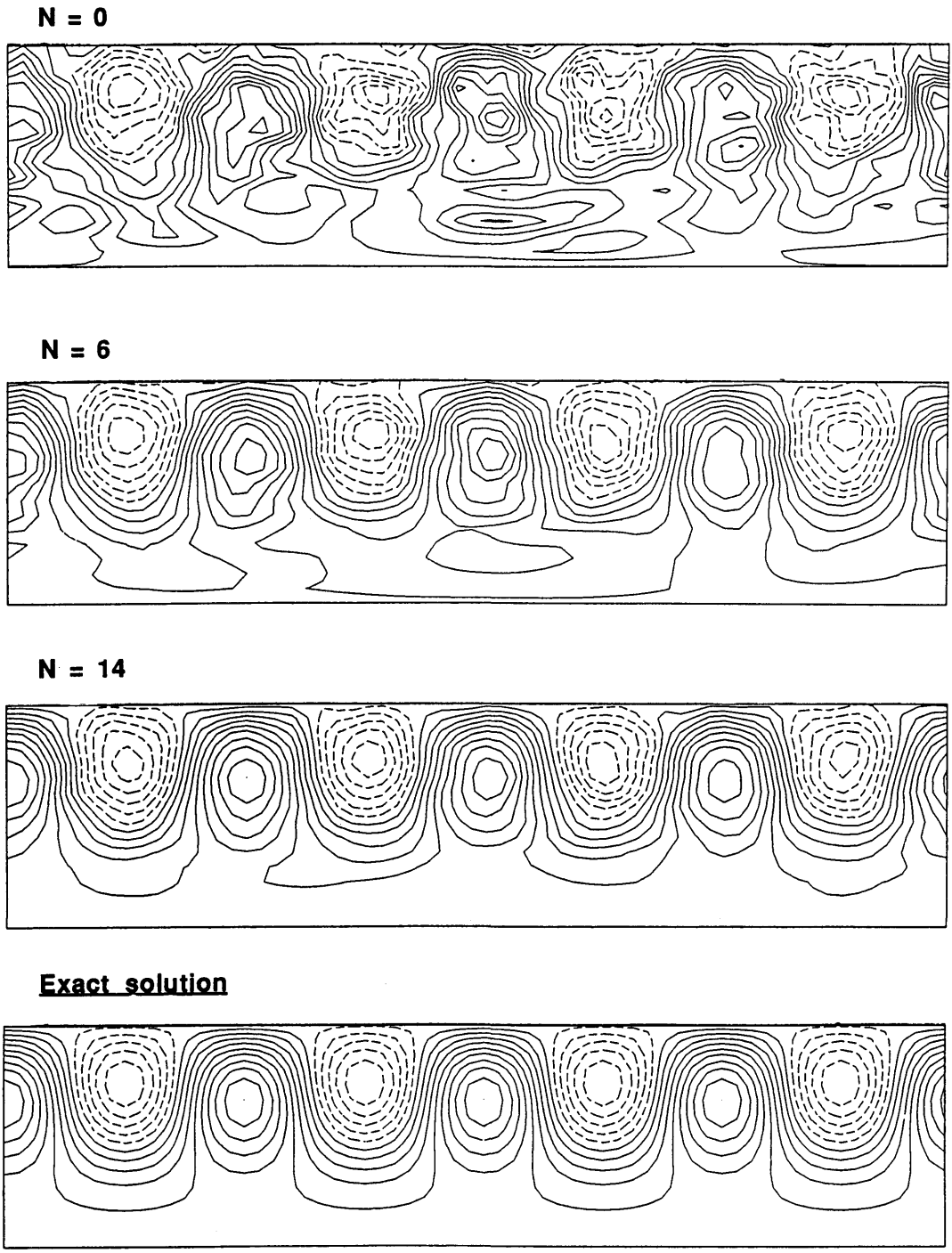

Fig. 13. Same as Fig. 12 but for the case where observations with a 5% relative error level are used: the field for $N = 0$ is representative of those observations. The last panel shows the configurations of the exact solution at $t = 0$.

with pseudo-random numbers ρ times 10^{-6} with $-0.5 < \rho < 0.5$ to fill all spectral components. By contrast, the first-guess used in TC was a state of rest. Fig. 12 shows the configuration of the vorticity field at $t = 0$ over the Northern hemisphere as the iteration went on. After 18 iterations, the solution was correctly recovered at which point $\|\nabla J\| < 0.1$ while initially, the norm of the gradient was bigger by three order of magnitudes.

In a second experiment, these same observations were contaminated with a random error field generated with pseudo-random numbers to create a 5% relative error level with respect to β, the initial amplitude of the Haurwitz wave. Fig. 13 shows the analyzed state over the Northern hemisphere as the iteration went on and the first guess ($N = 0$) corresponds to the observations used for $t = 0$. If one thinks of the observations as the best analysis available from a 3D scheme, these results show that more information can be extracted from their time evolution by imposing the constraint that they be dynamically consistent. Since it is not possible for the model to find initial conditions that would fit the random noise, the variational scheme extracts the part of the information that is dynamically coherent. However, the uncertainty on the gradient causes the convergence to occur at point Z_1^* where $\|\nabla J\|$ vanishes but the true gradient may not. Since the truth is available in this type of experiment, the true gradient was calculated: it has been found that, at Z_1^*, $\|\nabla J_t\| \sim 6.6$ while $\|\nabla J\| \sim 0.06$. So, had perfect observations been used, the iteration process would not have stopped at this point. A comparison of the final analyzed state against the exact solution (shown in the last panel of Fig. 13) shows that although the variational technique is able to improve upon the 3D analysis, the final analysis still differs from the truth.

4. Discussion and conclusion

The analysis and gradient error covariances can be related by making use of the tangent linear model. Since $Z(t_i) - Z_i^t = Q_i^T(Z_0 - Z_0^t)$, then

$$\nabla J(Z_0) = \frac{2}{N+1} \sum_{i=0}^{N} Q_i [Q_i^T (Z_0 - Z_0^t) - b_i]. \quad (13)$$

The analysis $Z_0 = Z_1^*$ being defined as the point where the gradient vanishes, (13) then implies that

$$\mathcal{M}(Z_1^* - Z_0^t) = \sum_{i=0}^{N} Q_i b_i$$

with

$$\mathcal{M} = \sum_{i=0}^{N} Q_i Q_i^T.$$

Each term of the sum that defines $\mathcal{M}$ is seen to correspond to a forward integration of the tangent linear model followed by a backward integration of the adjoint model. If $\mathcal{M}$ can be inverted, the analysis error is

$$Z' = Z_1^* - Z_0^t = \mathcal{M}^{-1} \sum_{i=0}^{N} Q_i b_i.$$

If the observational error is assumed to be uncorrelated in time as before, the analysis error covariance is shown to be

$$\overline{Z'Z'^T} = \frac{(N+1)^2}{4} \mathcal{M}^{-1} \overline{G'G'^T} \mathcal{M}^{-1}. \quad (14)$$

However for the Lorenz model, $\mathcal{M}$ cannot be inverted. This is a consequence of the fact that the action of this operator corresponds to a forward integration of the tangent linear model followed by a backward integration of the adjoint model. As can be seen from (6), the flow in phase space associated with both of these models has a constant divergence, negative for the tangent linear model and positive for the adjoint model. Therefore, a forward (backward) integration of the tangent linear model (adjoint model) will shrink a volume element to zero as $t \to \infty$: this property is related to the relaxation towards the attractor. Since $\mathcal{M}$ describes the combined action of these two integrations, it has to be singular. This can be verified on the two cases discussed in the preceding section. For both of them, the observation error covariance matrix is of the form $R_n = cI$, with I the identity matrix and c, a constant. It can then be shown from (10) that

$$\overline{G'G'^T} = \frac{4}{(N+1)^2} \mathcal{M}. \quad (15)$$

The determinants of the two matrices given by (11) and (12) are seen to be vanishingly small and as a consequence, $\mathcal{M}$ cannot be inverted. However, it is to be expected that $\mathcal{M}$ is invertible for the inviscid barotropic model considered in Section 3 since it preserves volume in phase space in which case (14) can be used to obtain the analysis error covariance. Finally, it should be mentioned that in Lacarra and Talagrand (1988), the linear operator $Q_i Q_i^{\mathrm{T}}$ (which is part of the definition of our $\mathcal{M}$) has been used to define the error amplification over a *finite* time interval.

The interest for quadri-dimensional data assimilation stems from its ability to extract the information contained in observations available over a finite time interval. The temporal aspect of the analysis raises new questions and the purpose of this paper was to illustrate some of these by using the Lorenz model. In Section 2, it has been shown that the presence of secondary minima can be a problem in some regions of phase space while not causing any serious difficulty in other regions. This shows that the performance of the method can be case dependent because the local error growth rate can vary from one situation to another (Benzi and Carnevale, 1989). This growth rate being more important for the case with transition, the same uncertainty in the analysis was seen to lead to a poorer forecast than what was observed in the regular case. (see Fig. 3 and 7). In Section 3, it has been shown how to relate the observation error covariance matrix to the gradient error covariance matrix and that the same error on the observations can have a different impact on the calculated gradient: this also is a consequence of the different error growth rates. Convergence of a minimization algorithm being determined by having a zero gradient, this incertitude in the gradient can be thought of as a measure of the accuracy of the final analysis and the gradient error covariance matrix provides an estimation of this uncertainty. Since it can be calculated for any model once its adjoint is known, it may prove to be a useful tool in more realistic systems to evaluate different scenarios proposed in observing systems simulation experiments.

Other aspects of data assimilation are worth looking at within the context of simple systems. In the present paper, all experiments were of the identical twin type, the model used to generate the observations being the same as the one employed to perform the assimilation. It could prove to be interesting to perform instead a distant cousin experiment for which the two models would be different: this is an important issue that is related to some aspects of the spin-up problem. This could be studied again with the Lorenz model by generating observations with a different parameter setting than the one used by the assimilating model.

5. Acknowledgments

The author would like to thank Professor Olivier Talagrand of the Laboratoire de Météorologie Dynamique in Paris for his careful review of the paper. Comments by Dr. Roger Daley of the Canadian Climate Centre, Dr. Philippe Courtier and Florence Rabier of Météo-France and an anonymous reviewer are also gratefully acknowledged. This research has been made possible through a Visiting Fellowship in Canadian Government Laboratories from the Natural Sciences and Engineering Research Council of Canada.

6. Appendix

Numerical integration scheme

As in Lorenz (1963), (5) is integrated with a second order Taylor scheme. With a timestep of $\Delta t = 0.001$, this scheme is sufficiently accurate not to lead to a significant error after an integration over 16 time units such as those considered in Section 2. If $Z = (X_1, X_2, X_3)^{\mathrm{T}}$ and using the short-hand notation of (1), it can be formulated as a predictor-corrector of the form

$$\tilde{Z}_{n+1} = Z_n + \Delta t F(Z), \tag{A.1a}$$

$$Z_{n+1} = Z_n + \frac{\Delta t}{2}(F(Z_n) + F(\tilde{Z}_{n+1})), \tag{A.1b}$$

with the subscript n referring to evaluation at time $t = n\,\Delta t$. This form is valid for any autonomous dynamical system.

For a non autonomous system such as the tangent linear model, the second order Taylor scheme is equivalent to

$$\delta\tilde{Z}_{n+1} = \delta Z_n + \Delta t A_n\,\delta Z_n,$$

$$\delta Z_{n+1} = \delta Z_n + \frac{\Delta t}{2}(A_n\,\delta\tilde{Z}_{n+1} + A_{n+1}\,\delta Z_n),$$

with $A(t)$ given by (6) and $A_n \equiv A(t_n)$. The linearity allows it to be put in the explicit form

$$\delta Z_{n+1} = q_n \, \delta Z_n,$$

with

$$q_n = I + \Delta t(A_n + A_{n+1}) + \tfrac{1}{2} \Delta t^2 A_n A_n,$$

I being the identity matrix. A solution of the tangent linear system can therefore be written as

$$\delta Z_i = Q_i^{\mathrm{T}} \, \delta Z_0,$$

with

$$Q_i^{\mathrm{T}} = q_{i-1} q_{i-2} \cdots q_0$$

being the discretized form of the resolvent. It then follows immediately that its adjoint is

$$Q_i = q_0^{\mathrm{T}} q_1^{\mathrm{T}} \cdots q_{i-1}^{\mathrm{T}}.$$

REFERENCES

Benzi, R. and Carnevale, G. F. 1989. A possible measure of local predictability. *J. Atmos. Sci. 46*, 3595–3598.

Courtier, P. and Talagrand, O. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Quart. J. R. Met. Soc. 113*, 1329–1347.

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

Derber, J. C. 1987. Variational four-dimensional analysis using quasi-geostrophic constraints. *Mon. Wea. Rev. 115*, 998–1008.

Ghil, M. 1989. Meteorological data assimilation for oceanographers. Part I: description and theoretical framework. *Dyn. Atmos. Oceans 13*, 171–218.

Guckenheimer, J. and Holmes, P. 1983. *Nonlinear oscillations, dynamical systems and bifurcations of vector fields*. Springer-Verlag, New York, 453 pp.

Hoffman, R. N. 1986. A four-dimensional analysis exactly satisfying equations of motion. *Mon. Wea. Rev. 114*, 388–397.

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

Le Dimet, F. X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus 38A*, 97–110.

Legras, B. and Ghil, M. 1985. Persistent anomalies, blocking and variations in atmospheric predictability. *J. Atmos. Sci. 42*, 433–471.

Lewis, J. M. and Derber, J. C. 1985. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus 37A*, 309–322.

Lorenc, A. C. 1988. Optimal nonlinear objective analysis. *Quart. J. R. Met. Soc. 114*, 205–240.

Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci. 20*, 130–141.

Lorenz, E. N. 1980. Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci. 37*, 1685–1699.

Lorenz, E. N. 1982. Atmospheric predictability with a large numerical model. *Tellus 34A*, 505–513.

Lorenz, E. N. 1990. Can chaos and intransitivity lead to interannual variability. *Tellus 42A*, 378–389.

Miller, R. N. and Ghil, M. 1990. Data assimilation in strongly nonlinear current systems. In: *Proceedings of the WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography*. Clermont-Ferrand, France, July 9–13, 1990. pp. 93–98.

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

Pedlosky, J. and Frenzen, C. 1980. Chaotic behavior of finite-amplitude baroclinic waves. *J. Atmos. Sci. 37*, 1177–1196.

Saltzmann, B. 1962. Finite amplitude free convection as an initial value problem. *J. Atmos. Sci. 19*, 329–341.

Talagrand, O. and Courtier, P. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I. Theory. *Quart. J. R. Met. Soc. 113*, 1311–1328.

Thompson, P. D. 1969. Reduction of analysis error through constraints of dynamical consistency. *J. Atmos. Sci. 8*, 738–742.