Tellus (1996), 48A, 96–121
Printed in Belgium – all rights reserved

TELLUS
ISSN 0280–6495

On extending the limits of variational assimilation in nonlinear chaotic systems

By CARLOS PIRES,* ROBERT VAUTARD and OLIVIER TALAGRAND, *Laboratoire de Météorologie Dynamique du CNRS, Paris, France*

(Manuscript received 21 September 1994; in final form 10 March 1995)

ABSTRACT

A study is made of the limits imposed on variational assimilation of observations by the chaotic character of the atmospheric flow. The primary goal of the study is to determine to which degree, and how, the knowledge of past noisy observations can improve the knowledge of the present state of a chaotic system. The study is made under the hypothesis of a perfect model. Theoretical results are illustrated by numerical experiments performed with the classical three-variable system introduced by Lorenz. Both theoretical and numerical results show that, even in the chaotic regime, appropriate use of past observations improves the accuracy on the estimate of the present state of the flow. However, the resulting estimation error mostly projects onto the unstable modes of the system, and the corresponding gain in predictability is limited. Theoretical considerations provide explicit estimates of the statistics of the assimilation error. The error depends on the state of the flow over the assimilation period. It is largest when there has been a period of strong instability in the very recent past. In the limit of infinitely long assimilation periods, the behaviour of the cost-function of variational assimilation is singular: it tends to fold into deep narrow "valleys" parallel to the sheets of the unstable manifold of the system. An unbounded number of secondary minima appear, where solutions of minimization algorithms can be trapped. The absolute minimum of the cost-function always lies on the sheet of the unstable manifold containing the exact state of the flow. But the error along the unstable manifold saturates to a finite value, and the absolute minimum of the cost function does not, in general, converge to the exact state of the flow. Even so, the absolute minimum of the cost function is the best estimate that can be obtained of the state of the flow. An algorithm is proposed, the *quasi-static variational assimilation*, for determining the absolute minimum, based on successive small increments of the assimilation period and quasi-static adjustments of the minimizing solution. Finally, the impact of assimilation on predictability is assessed by forecast experiments with that system. The ability of the present paper lies mainly in the qualitative results it presents. Qualitative estimates relevant for the atmosphere call for further studies.

1. Introduction

It is now well established that the quality of weather forecasts is highly dependent on the quality of the initial conditions. This sensitivity imposes an ultimate limit to deterministic weather prediction, estimated to be about two weeks, which no conceivable observing and forecasting system will ever allow to cross. But present performance of numerical weather prediction shows that this ultimate limit is far from being reached, and that substantial gain can still be obtained, both in the quality of short-range forecasts and in the extension of the range of useful forecasts, through more accurate specification of the initial conditions. This can be obtained, not only through improvement of the observing system itself, but also through improvement of the complex sequence of operations which, starting from the

*Corresponding author.

raw observations, leads to the initial conditions of the forecasts. *Assimilation of meteorological observations* is that sequence of operations.

The accurate determination of the current state of the flow usually requires the use of observations distributed over a recent-past period. In fact, the available relevant information consists not only of the observations proper, but also of the physical laws which govern the temporal evolution of the flow, which are generally available under the form of a numerical model, and assimilation of observations is an appropriate combination of the two information sources. It must take into account their uncertainty, since neither source can be expected to be exact. In addition to an estimate of the state of the flow, it must also produce an estimate of the resulting estimation error.

Assimilation of observations, originating from the needs of numerical prediction is now extending to other domains of application. First, the rapid progress in dynamical oceanography, particularly in general circulation models development and new observing systems, has stimulated the assimilation of oceanographic data (see, e.g., Ghil and Malanotte-Rizzoli, 1991, and references therein). Secondly, the development of climatic studies makes it desirable to build as accurate and as homogeneous as possible descriptions of the atmospheric flow over long periods of time. To that end, several major meteorological centres are involved in the "reanalysis" of past data. There exists a fundamental difference between assimilation performed for the purpose of weather prediction or for the purpose of reanalysis of past data. In the former case, the model is to be run into the future from the initial conditions produced by the assimilation of past observations. In the latter case, observations performed both before and after estimation time are available, and it is a priori preferable to develop algorithms capable of using all these observations.

A detailed description of the state of development of the theory of assimilation, and of the associated numerical algorithms, has been given by Daley (1991) (for other basic references on assimilation, see Lorenc, 1986; Ghil and Malanotte-Rizzoli, 1991; Talagrand, 1992; Bennett, 1992). Even though the atmosphere and the ocean are highly nonlinear—actually chaotic—systems, most, if not all, numerical algorithms that have been used either for operational assimilation or for research purposes can be described as more or less simplified forms of statistical linear estimation. Statistical linear estimation (Jazwinski, 1970) determines the *Best Linear Unbiased Estimator*, or *BLUE*, of the state of the flow, i.e., the linear combination of the observations (in the broad sense of all the known quantities from which the estimation is to be made) that minimizes the variance of the corresponding estimation error. The explicit determination of the BLUE requires the prior knowledge of the first- and second-order statistical moments of the errors affecting the various sources of information, i.e., the observations proper on the one hand and the assimilating model on the other. Mathematically, the BLUE can also be determined as the minimizer of the sum of the squared observations-minus-estimated-values differences, weighted by the inverse of the observation error covariance matrix.

From a practical point of view, two broad classes of algorithms have been defined for assimilation: *sequential and variational* algorithms. In sequential algorithms, the model is integrated forward in time over the assimilation period. Whenever the model time reaches an instant at which observations are available, the state predicted by the model, usually called the *first-guess* in that context, is corrected with the new observations. The corrected state is obtained as a linear combination of the first-guess and the observations, weighted according to their respective assumed accuracies. All assimilation algorithms which have been used so far in operational numerical weather prediction are of the sequential type. They can be described as more or less simplified forms of *Kalman filtering* (Kalman, 1960; Jazwinski, 1970; Ghil and Malanotte-Rizzoli, 1991). At any stage, Kalman filtering produces the BLUE of the state of the system, taking into account all observations performed before this stage. However, the numerical cost of exact Kalman filtering, in a meteorological or oceanographical context, and with present dynamical models and computers, would be totally prohibitive. That is, among others, one very good reason why present operational algorithms for assimilation, often of the form called *optimal interpolation* (Lorenc, 1981), are only simplified forms of Kalman filtering.

Variational algorithms determine a nonlinear best estimate of the flow through direct explicit minimization of the scalar function measuring the

misfit between the observations and the estimated values. If, as will be done in this paper, the model is assumed to be perfect, the misfit function, or *cost function*, to be minimized is the sum of the squared observation-minus-model differences, weighted by the inverse covariance matrix of the observational errors. Variational assimilation globally adjusts a model solution to the observations available over the entire assimilation period. It thus ensures that the information contained in the observations is propagated both forward and backward in time, and the assimilated state at any instant within the assimilation period depends on the observations performed over the entire period. In the case of a linear system, the sequence of state produced by variational assimilation is identical with the sequence of states that would be obtained from Kalman smoothing (Bennett and Budgell, 1989; Gaspar and Wunsch, 1989), an extension of Kalman filtering taking into account all observations over the assimilation period.

From a practical point of view, variational assimilation is implemented through iterative minimization with respect to the state of the model at the initial time of the assimilation period. This is possible through the use of the *adjoint* equations of the model, which allow to explicitly compute, at a non-prohibitive computational cost, the gradient of the cost-function with respect to the state of the model at the beginning of the assimilation period (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987). It has also been found profitable to add to the cost function a *background term* containing the estimation of the flow at the initial time of the assimilation period, obtained from the previous assimilation cycles. Under these conditions, variational assimilation is much more economical than Kalman filtering. Numerous studies have now been performed on variational assimilation, in both the meteorological and oceanographical contexts (Lewis and Derber, 1985; Thacker and Long, 1988; Courtier and Talagrand, 1990; Sheinbaum and Anderson, 1990a and b; Zou et al., 1992; Thépaut et al., 1993; see also the references in Courtier et al., 1993). The results are promising enough so that several major meteorological centers have decided to develop variational assimilation for operational purposes (Courtier et al., 1994).

The theory of BLUE, being a linear regression in essence, assumes that the model equations be linear, and that the observations proper be linear functions of the model variables. In practice, information is available that is much more accurate than relationships based on statistical linear regression. The models of the dynamics of the atmospheric or oceanic flows, built on the physical principles of conservation of mass, energy and momentum, are indeed nonlinear. A similar remark applies, for instance, to the infra-red radiances measured by satellites, which are related, through the strongly nonlinear (and explicitly known) radiative transfer equation, to the corresponding atmospheric temperature and humidity fields. The question is therefore whether it is legitimate to use linear algorithms in order to solve intrinsically nonlinear estimation problems. The important point is that assimilation methods always use a first approximation of the state of the flow to be estimated, under the form of a "first-guess". What must in effect be estimated is the *deviation* of the state of the flow from that first approximation. The cost function used in variational assimilation can be formulated in terms of this departure by using the *tangent linear* approximation, in which case the BLUE theory is also valid. In that case only, the cost-function is quadratic, with a unique minimum that can be determined through standard minimization algorithms. Within the tangent linear approximation, an appropriate sequential algorithm is the *extended Kalman filter* (Budgell, 1986).

In the meteorological context, the validity of the tangent linear approximation has been explicitly verified by various authors, for large-scale flows and assimilation periods of two or three days (Lacarra and Talagrand, 1988; Rabier and Courtier, 1992), and for mesoscale flows, with shorter periods of the order of 36 hours (Errico and Vukicevic, 1992). This difference is due to the existence of strongly nonlinear processes such as convection (Parrish and Derber, 1992; Courtier et al., 1994). The validity of the linear tangent approximation in the assimilation of oceanographic data has apparently not been checked so far. Budgell (1986) has found that the extended Kalman filter, used in simplified but realistic conditions, was able to properly assimilate synthetic data, thus supporting the validity of the tangent linear hypothesis. On the other hand, Evensen (1992), also in simplified (but no less realistic) conditions, has observed numerical divergence of the

extended Kalman filter, as confirmed by later tests (Evensen, 1994), that divergence was due to neglect of nonlinear saturation processes. In any case, the tangent linear approximation becomes invalid when the assimilation period is extended over long periods, typically more than 3 days for large scale meteorological flow. For long assimilation periods, the chaotic character of the flow will progressively become more and more important, and it is necessary to determine the resulting implications for assimilation. The purpose of our article is precisely to determine the impact of taking long assimilation periods.

Assimilation with nonlinear chaotic systems has already been studied by several authors. Gauthier (1992), Stensrud and Bao (1992), and Miller et al. (1994) have performed experiments with the classical three-parameter system introduced by Lorenz (1963), using either sequential or variational algorithms. All these authors have found that performance of assimilation varies significantly depending on the period of time over which it is implemented, and that assimilation is particularly difficult at times when the system experiences "transitions" between the two lobes of the Lorenz attractor. They have also observed that extension of variational algorithms over long periods raises difficulties due to the occurrence of multiple minima in the cost-function. In the presence of multiple minima, the result of the minimization will depend on the starting point of the minimization.

The present paper is a further step in the study of assimilation with chaotic systems. Our goal is to investigate quantitatively the performance of fully-nonlinear variational assimilation, and to determine its theoretical limits, as well as the ensuing implications for deterministic predictability. We shall consider only a highly idealized situation, where

(i) The model is perfect, that is, we know the exact evolution equations of the dynamical system under observation. This hypothesis is made for the sake of simplicity. This simplification will be discussed in Section 6.

(ii) Observations of the complete state of the system are available at regularly spaced intervals of time. These observations are contaminated by observational noise, but are available over arbitrary long periods in the past.

The main question we address is: what accuracy can be obtained through variational assimilation, from the information contained in (i) and (ii), on the state of the system and on its subsequent evolution? This question raises a number of subsidiary points. For instance, how does the accuracy of the assimilated initial state vary when one uses progressively longer assimilation periods, eventually going back to infinity? In other words, what is the information content of remote past observations? Finally, how is it practically possible to find the absolute minimum of the cost function and what is its relevance relative to other minima?

Since our goal is primarily theoretical, we shall illustrate our purpose on a low-dimensional dynamical system: the already-mentioned classical three-dimensional differential system of Lorenz (1963) and the two-dimensional mapping introduced by Hénon (1976). Both systems, like the atmosphere, are dissipative and chaotic, and therefore possess a "strange attractor" to which the orbits converge over long periods. The low dimensionality of these systems allows not only easy numerical experimentation, but also relatively easy "understanding" of the chaotic character of the corresponding dynamics. It must be emphasized however, that the applicability of the results presented below to more complicated systems as general circulation models requires further testing.

In Section 2, we study qualitatively on several examples the behaviour of the cost-function when the length of the assimilation period increases to infinity, and we suggest how the corresponding absolute minimum can be tracked by progressively increasing, through successive small increments, the length of the assimilation period. In Section 3 we present the statistical properties of the absolute minimum, and of the covariance matrix of the associated estimation error. We estimate also the past time beyond which the information content of observations becomes negligible. We develop in Section 4 the quasi-static variational assimilation (QSVA), through which the absolute minimum of the cost-function can be tracked over assimilation periods of increasing length. It is applied to the Lorenz system. Section 5 is devoted to the quantification of the predictability gain brought by assimilation over long periods. Conclusions follow in Section 6.

2. Behaviour of the cost function in chaotic systems

2.1. *The stable manifold, the unstable manifold and the cost function*

We consider a dynamical system described by a state vector $x(t)$ whose time evolution is given by a finite set of m autonomous differential equations of the form

$$\frac{dx}{dt} = F(x). \tag{2.1}$$

Unless specified otherwise, we assume for the sake of simplicity that observations $\tilde{x}(t_i)$ have been taken at a constant sampling rate δt ($t_i = -i\,\delta t$). The observation error, $\varepsilon(t_i) = \tilde{x}(t_i) - x(t_i)$, is assumed to be random, uncorrelated in time, and independent of both the time of the observation and the observed state $x(t_i)$. Without loss of generality, we can assume that the error is isotropic, with variance σ^2 along any direction. This hypothesis may require a prior linear transformation on the state vector, which does not change any of the topological properties of the dynamical system, but renders the calculations much easier.

We consider assimilations performed on the latest $N+1$ observations, that is, over an assimilation period of length $\tau = N\,\delta t$. In general, the *final* time t_0 of the assimilation period will be kept the same (and will sometimes be called, in reference to the situation where one wants to predict the future evolution of the system, the *present* time), while the *initial* time t_N will vary. In particular, we will consider the behaviour of the assimilation in the limit $N \to \infty$.

Given a solution $\hat{x}(t)$ of (2.1), and given our assumptions on the noise, the cost-function reads

$$J(\tau, \hat{x}(0), x(0)) = \frac{1}{N+1} \sum_{i=0}^{N} \|\hat{x}(t_i) - x(t_i) - \varepsilon(t_i)\|^2. \tag{2.2}$$

where $\|\cdot\|$ denotes the norm associated to a scalar product denoted $\langle \cdot, \cdot \rangle$ in the following. J measures the misfit between the solution $\hat{x}(t)$ and the observations $\tilde{x}(t_i) = x(t_i) + \varepsilon(t_i)$. The purpose of variational assimilation is to determine the solution of (2.1) which minimizes J. In (2.2), J is considered as a function of the *estimated* state vector $\hat{x}(0)$ (denoted hereafter $\hat{x}$) at the present time, of the *real* state vector $x(0)$ (denoted hereafter x) at the same time, and of the length τ of the assimilation period. Since the system is assumed to be entirely deterministic, the cost function can alternatively be expressed in terms of the real and estimated state vectors at the initial time t_N of the assimilation period, rather than at the final time,

$$J'(\tau, \hat{x}(t_N), x(t_N)) = J(\tau, \hat{x}, x). \tag{2.3}$$

The cost-function J defined by (2.2) will be called the *backward* cost-function, while the cost-function J' defined by (2.3) will be called the *forward* cost-function.

In practice, explicit computation of the backward cost function would require backward integrations of the model. For a system which contains dissipative processes, numerical errors are strongly amplified by dissipation in a backward integration, leading to rapid numerical divergence. Therefore, variational assimilation is implemented in practice on the forward cost function J', and the minimization is performed on the estimated state vector $\hat{x}(t_N)$ at the beginning of the assimilation period. It will nevertheless be convenient for the theoretical developments which will follow to consider the backward cost-function (2.2).

Eq. (2.2) can be decomposed into

$$\begin{aligned} J(\tau, \hat{x}, x) = {} & \frac{1}{N+1} \sum_{i=0}^{N} \|\hat{x}(t_i) - x(t_i)\|^2 \\ & + \frac{1}{N+1} \sum_{i=0}^{N} \|\varepsilon(t_i)\|^2 \\ & - \frac{2}{N+1} \sum_{i=0}^{N} \langle \hat{x}(t_i) - x(t_i), \varepsilon(t_i) \rangle. \end{aligned} \tag{2.4}$$

The first term on the right-hand side of (2.4) is the *error-free* cost-function $J_e(\tau, \hat{x}, x)$, that we would obtain in the absence of observational error. The second term is the average of the squared observational error over the assimilation period, and the third term is a cross-product between the difference $\hat{x}(t_i) - x(t_i)$ and the observational error.

When τ goes to infinity, the second term converges to the total variance of the observation error, $E^2 = m\sigma^2$, independently of the two model trajectories $x(t)$ and $\hat{x}(t)$. In the same limit, the last term goes to zero if we make the assumption that there is no correlation between the observation

error and the dynamical variables, since $\hat{x}(t_i) - x(t_i)$ is bounded. For large values of N, the variations of J with $\hat{x}$ will be dominated by the variations of J_e. In particular, for long assimilation periods, low values of J should coincide with low values of J_e.

J_e measures the mean square distance, accumulated over the assimilation period, between the two model solutions $x(t)$ and $\hat{x}(t)$. Two model solutions cannot generally remain close to each other in the remote past, even if they are very close at time 0, owing to the fact that in chaotic dissipative systems, sensitivity to initial conditions acts both forward and backward in time. Hence, for most trajectories, the cost function J_e tends, as τ goes to infinity, to a non-zero or infinite limit. However, some trajectories $\hat{x}(t)$ converge, as $t \to -\infty$, to the real trajectory $x(t)$. These are the trajectories originating at time 0 from the *unstable manifold* of $x = x(0)$ (Guckenheimer and Holmes, 1983).

The *unstable manifold* $W_u(x)$ of x is defined as the set of state vectors $\hat{x} = \hat{x}(0)$ of the phase space verifying the condition:

$$\hat{x} \in W_u(x) \Leftrightarrow \lim_{t \to -\infty} \|\hat{x}(t) - x(t)\| = 0. \tag{2.5a}$$

The *stable manifold* $W_s(x)$ is defined in a similar manner:

$$\hat{x} \in W_s(x) \Leftrightarrow \lim_{t \to +\infty} \|\hat{x}(t) - x(t)\| = 0. \tag{2.5b}$$

It has been proven that these manifolds exist (Hirsch and Pugh, 1970) for a special class of systems, the hyperbolic systems. In fact, for such systems, there are constants $C > 0$ and η $(0 < \eta < 1)$ such that, when x lies on the attractor of the system,

$$\|\hat{x}(t) - x(t)\| \leqslant C\eta^{|t|}, \quad \text{for} \quad \hat{x} \in W_u(x) \quad \text{and } t < 0, \tag{2.6a}$$

$$\|\hat{x}(t) - x(t)\| \leqslant C\eta^{|t|}, \quad \text{for} \quad \hat{x} \in W_s(x) \quad \text{and } t > 0. \tag{2.6b}$$

Since $|\eta| < 1$, the cost-function $J_e(\tau, \hat{x}, x)$ converges exponentially to zero, when τ goes to infinity, if $\hat{x}$ lies on the unstable manifold. Thus, the limit cost function J_e' is somewhat singular: it vanishes along the unstable manifold, while it is constant elsewhere, equal to the average square distance between two random points in the attractor. In the presence of observation errors, the limit cost function is also singular: it reaches its minimal value, E^2, everywhere along the unstable manifold. For finite values of τ, however, the absolute minimum of J_e is always obtained for $\hat{x} = x \equiv x(0)$. Intuitively, one easily understands that the point in phase space corresponding to the absolute minimum of the complete cost function J (2.4) must rapidly converge, for large τ, to the unstable manifold. Along this manifold, the fluctuations of the last two terms in (2.4) dominate the fluctuations of J. As will be shown in Section 3, the point in phase space corresponding to the absolute minimum does not generally converge to the exact initial condition x, but to a point lying on the unstable manifold.

A dual situation occurs for the forward cost function, J', the stable manifold simply playing the role played by the unstable manifold in the backward mode. When the assimilation period is extended both backward and forward in time, the point in phase space corresponding to the absolute minimum of the cost function converges to the exact vector state x. This may have useful consequences for the reanalysis of past observations. For the same length of the assimilation period, and for the same accuracy and spatio-temporal distribution of observations, the state of the atmosphere at time t should be reconstructed much more accurately if one considers an assimilation period containing past *and* future observations than only past observations. This is in agreement with the fact that assimilation error is minimal inside the assimilation period (Derber, 1987). In schematic terms, past observations provide accuracy about the present state along its stable modes, while future observations provide accuracy along its unstable modes.

These results are illustrated by Fig. 1, for to the two-dimensional discrete mapping introduced by Hénon (1976)

$$x_{n+1} = 1 + y_n - ax_n^2, \tag{2.7a}$$

$$y_{n+1} = bx_n, \tag{2.7b}$$

where x_n and y_n are the coordinates of the nth iterate of the mapping (x here denotes here a scalar component, and no more a complete state vector). Fig. 1 has been obtained for the values $a = 1.4$ and $b = 0.3$, for which system (2.7) is chaotic and dis-

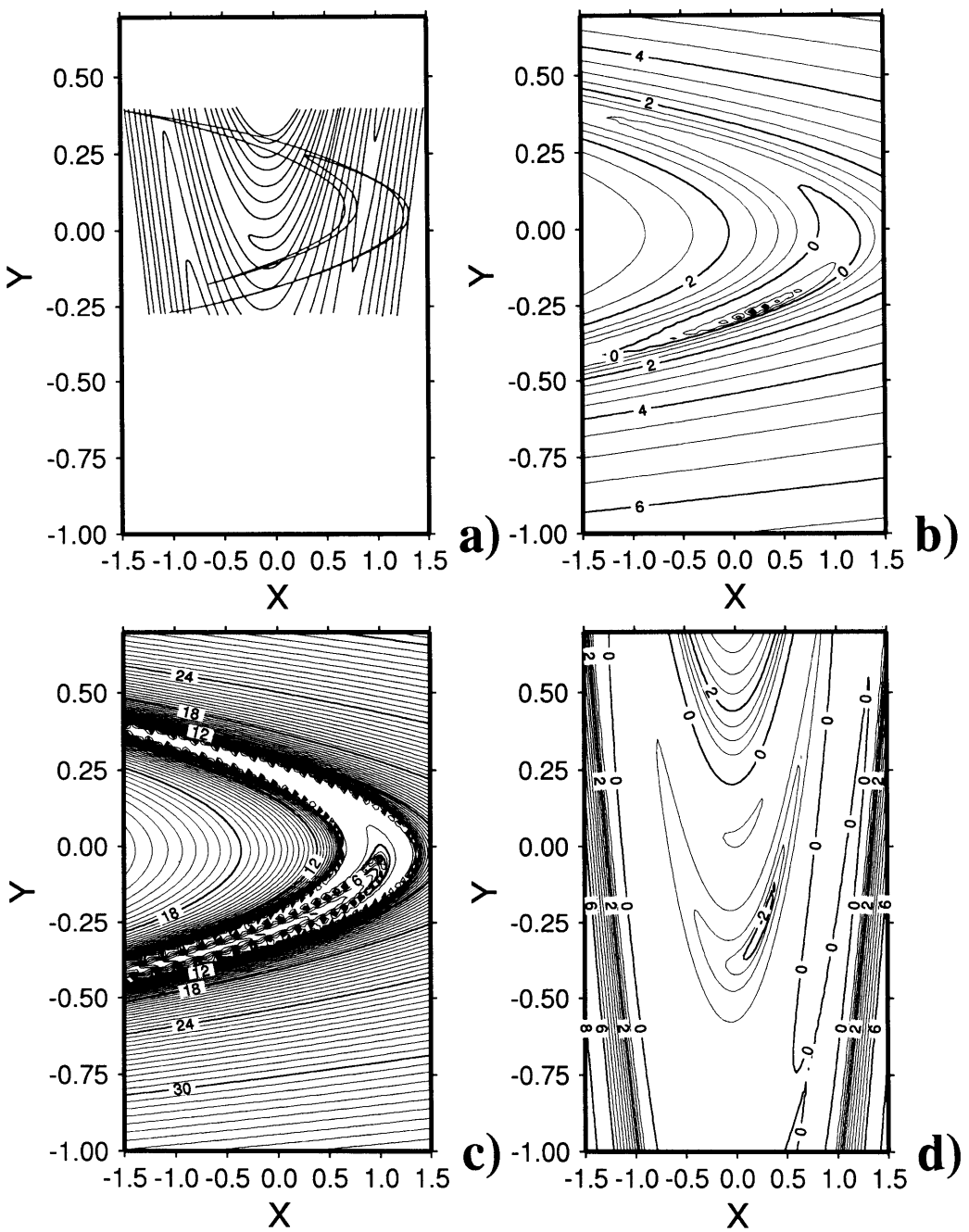

Fig. 1. Panel (a). Schematic representation of the stable and unstable manifolds of the Hénon mapping at point P (0.25; -0.25) (after Ruelle, 1989). See details in the text. Panels (b) and (c). Variations of the error-free backward cost function J_e, for assimilation periods of length $\tau = 2$ and 4, respectively. The cost-function is relative to an orbit going through the same point P as in panel (a). The contour lines are in decimal logarithms. Panel (d). Same as in panels (b) and (c), but for the error-free forward cost function J'_e (and $\tau = 4$).

sipative. Panel (a), borrowed from Ruelle (1989), is a schematic representation of the stable and unstable manifolds of the point P (0.25, −0.25). The unstable manifold is the crescent-shaped structure in the middle of the panel. It is entirely contained in the region of the (x, y)-plane shown in the panel. The stable manifold of P is the multiply folded curve. It extends to infinity, and only part of it is represented in the panel (the difference between the spatial extensions of the two manifolds is due to the dissipative character of the mapping). The representation is schematic in that both manifolds are not ordinary curves, but fractal sets consisting of an infinite number of foliations.

Panels (b) and (c) of Fig. 1 represent the variations, as a function of the estimated state at time 0 (the analog of $\hat{x}(0)$ of eq. (2.2)), of the error-free backward cost function J_e. The real state of the system at time 0 is the same point P as before, and panels (b) and (c) correspond to the respective values $N=2$ and 4 (here $\delta t=1$). In both panels, regions of low cost values stretch along the foldings of the unstable manifold. The stretching is much more pronounced for $N=4$ (panel 1(c)). In the vicinity of point P, J_e varies much more slowly along the direction of the unstable manifold $W_u(P)$ than along the orthogonal direction. Panel (d) of Fig. 1 represents the variations of the forward cost function J'_e, for the same point P and for $N=4$. It is now along the direction of the stable manifold $W_s(P)$ that the cost function varies slowly. We verified that, as N goes to infinity, and for both the forward and the backward cost functions, the isocontours experience repeated foldings which can be described by a binary scheme. This is a consequence of the existence in the Hénon map of an invariant set called a Smale horseshoe (Devaney and Nitecki, 1979). This situation shows a close relationship between the shape of the cost-function and the *symbolic dynamics* of the system (see Subsection 2.2 below). This folding mechanism is responsible for the occurrence of multiple minima, as observed from the experiments performed with the Lorenz system by Gauthier (1992) and Stensrud and Bao (1993).

2.2. *The foldings of the cost function and the symbolic dynamics*

We show now some other important features of the cost function using the well known Lorenz system (1963). The advantage of using this system as a prototype for atmospheric studies is that, in addition to being chaotic and dissipative, it shares with the atmospheric system some dynamical features, such as the existence of multiple "regimes", called *weather regimes* in the atmospheric context (Legras and Ghil, 1985; Vautard, 1990), separated by abrupt transitions during which predictability can be poor (Tibaldi and Molteni (1990); Palmer (1993)). The Lorenz system is described by a set of differential nonlinear coupled equations in the three variables: $(x, y, z) \in R^3$:

$$\frac{dx}{dt} = \sigma(y - x), \tag{2.8a}$$

$$\frac{dy}{dt} = \rho x - y - xz, \tag{2.8b}$$

$$\frac{dz}{dt} = -\beta z + xy. \tag{2.8c}$$

All the experiments discussed below are based on the reference trajectory presented in Fig. 2, which shows the time evolution of the variable $x(t)$ in some time interval $(-15, 25)$, for the classical parameter values $\sigma=10$, $\rho=28$, $\beta=8/3$ leading to a chaotic attractor (Sparrow, 1982). The system is integrated using the predictor-corrector scheme with a time-step of 0.02 units, starting from the initial conditions $(-4.62, -6.61, 17.94)$ at $t=-15$ (lying in the attractor, chosen after a long numerical integration). In all experiments described in this paper, synthetic "observations" are constructed from the reference trajectory of Fig. 2 by addition of random noise. Assimilation experiments are designed to reconstruct the reference trajectory from the noisy observations.

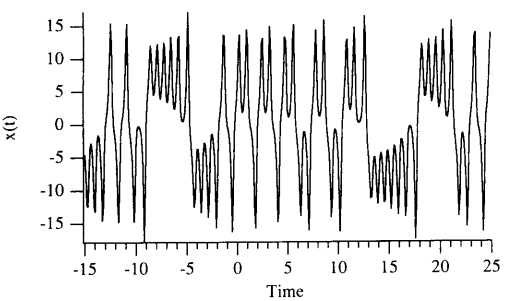

Fig. 2. Time variations, along the reference solution, of the variable $x(t)$ of the Lorenz system.

The two regimes of the Lorenz system correspond to oscillations around either one of two unstable fixed points of the system, which are symmetrical about the axis $x = y = 0$. The oscillations around one fixed point, during which the variable x keeps the same sign, are clearly visible in Fig. 2. They have a period of typically 1 time unit, and their amplitude progressively increases until, after a few oscillations, a transition occurs to the other fixed point. The transitions correspond to changes in the sign of x. Since the Lorenz system consists in a set of three differential equations, it has three Lyapunov exponents: one is positive, $\Lambda_1 \cong 1$, another is zero, $\Lambda_2 = 0$, and the third one is negative, $\Lambda_3 \simeq -14$ (Sparrow, 1982; Nese et al., 1987). The dimension of both the local stable and unstable manifolds is therefore 1, and there is a *neutral* manifold, spanned locally by the instan-

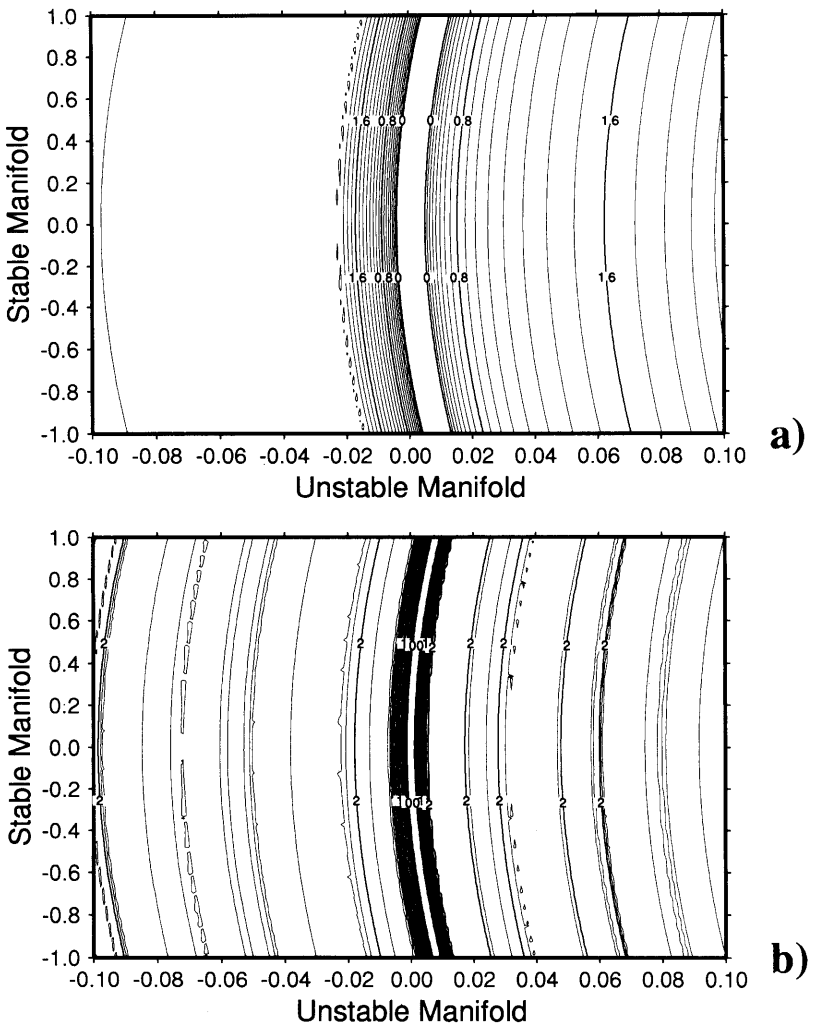

Fig. 3. Variations of the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ (Lorenz system) in the plane spanned by the stable and unstable directions, as determined from the tangent linear system (see text), and for $\tau = 6$ (panel (a)) and $\tau = 8$ (panel (b)) respectively. The metric has been distorted in order to make the stable and unstable manifolds orthogonal to each other in the figure. The scale on the contour lines is logarithmic (decimal logarithm). Contour interval: 0.1. For clarity, negative contours, which would be present only in the central "valley" directed along the stable manifold, have not been drawn.

taneous time-derivative vector (dx/dt, dy/dt, dz/dt). In practice, this neutral manifold disappears with time discretization, but the associated exponent remains very small. Errors diverge exponentially forward in time along the unstable manifold with an e-folding time $1/\Lambda_1$ (for infinitesimal errors and infinite time) and converge to zero along the stable manifold with an e-folding time $-1/\Lambda_3$ (Ruelle, 1989).

We shall hereafter study the behaviour of the cost function in a plane tangent to the local stable and unstable manifolds. The direction of the unstable manifold at time $t=0$ can be determined by simply integrating the *tangent linear system* (see Section 3 below) forward from some negative time $t=-T_u$, from a randomly chosen perturbation and renormalizing the vector at each regular short time intervals. The unstable direction is the direction of the resulting vector at time $t=0$. In a similar manner, the direction of the stable manifold at $t=0$ is defined by backward integration of the tangent linear system from a positive time $t=+T_s$. Here, we take $T_u=2$ and $T_s=0.4$. We focus here on the error-free forward cost function J'_e, keeping in mind the duality between unstable and stable manifolds, since long backward integrations of the dissipative system are numerically impossible.

Fig. 3, built in the same manner as Figs. 4 and 8 of Gauthier (1992), shows the variations of $J'_e(\tau, \hat{x}, x)$, for $\tau=6$ and $\tau=8$ (panels (a) and (b) respectively), and for an observation sampling rate $\delta t=0.1$. The variations of J'_e are represented in the neighbourhood of the state vector at time $t=0$ (see Fig. 2), corresponding to the approximate values $(x, y, z)=(6.55; 10.60; 16.62)$, and chosen as the origin in Fig. 3. Note that in both panels, in order to resolve the fine structure of the cost-function, the coordinate along the unstable direction (the horizontal coordinate) is magnified by a factor 10 with respect to the coordinate along the stable direction. The contours of J'_e tend to stretch along the stable manifold. The variations of J'_e in the unstable direction become sharper and sharper as τ increases. This is associated with foliations of the isocontours of J'_e along the stable direction, and the occurrence of a large number of secondary minima, visible in panel (b).

Fig. 4a shows cross-sections of the cost function J'_e along the horizontal axis of Fig. 3 (the unstable direction), for $\tau=8$, 9 and 9.7 (note the magnification of the horizontal axis in comparison with Fig. 3). For $\tau=8$, the cost function exhibits a parabolic shape centered at the origin. For $\tau=9$, the parabola has changed into a pair of two new parabolas separated by a cusp. This changing pattern is repeated in a self-similar manner when τ grows from 9 to 9.7.

The solid curve of Fig. 4b, which also corresponds to $\tau=9.7$, shows the variations of J'_e along the unstable direction (the dashed curve will be discussed later). The number of secondary minima has become very large (note that the range of variation along the unstable direction is ten times as large in Fig. 4b as in Fig. 4a). In order to understand better the creation of cusps and secondary minima as τ increases, we display next in Fig. 5 the time evolution of the coordinate x for the orbits originating from the absolute minimum, P (the origin), and from the three local minima A, B and C indicated in Fig. 4b. The trajectory originating from A, the secondary minimum closest to P, sticks to the reference trajectory originating from P

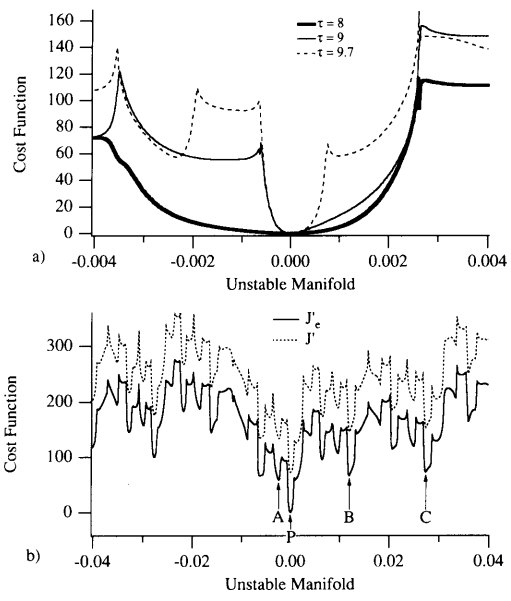

Fig. 4. Panel (a): Cross-section of the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ along the unstable manifold, for various values of τ. Panel (b). As in panel (a), for $\tau=9.7$, and with a display interval ten times as large, respectively for the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ (solid curve) and for the error-contaminated cost-function $J_e(\tau, \hat{x}, x)$ (dashed curve). In the latter case, the total variance of the observational noise is $E^2=75$.

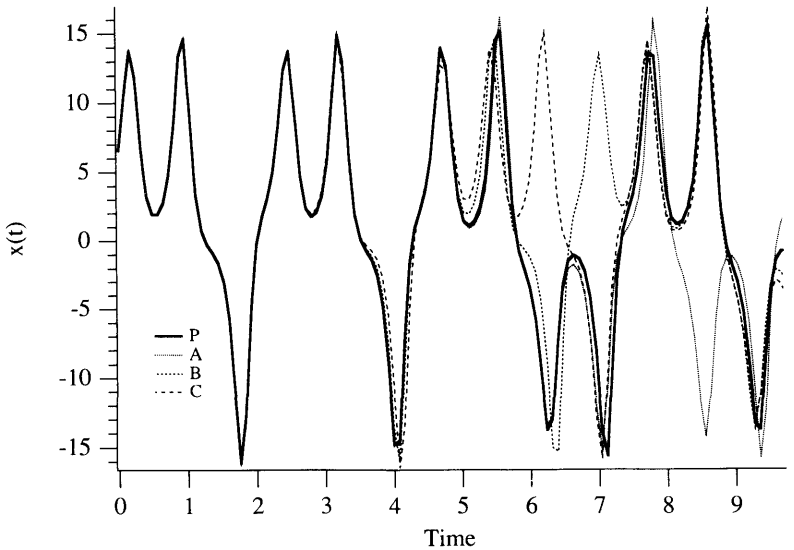

Fig. 5. Variations of the coordinate x along the orbits originating from the minima *P*, *A*, *B*, *C* (indicated in Fig. 4b) of the error-free cost-function.

up to about 8 time units, and thereafter undergoes an "erroneous" regime transition. The orbit emanating from the next closest secondary minimum *B* fails the transition near $t = 6.5$, while the orbit emanating from the farthest local minimum *C* misses a regime transition of the reference orbit at $t = 5.8$. In all three cases, the perturbed orbits shift back to the reference orbit one oscillation after they have left it.

These results suggest a simple interpretation for the fluctuations of the cost-function. Each secondary minimum corresponds to an orbit which follows the same sequence of oscillations as the reference orbit around one of the two fixed points of the system, before missing a transition (or undergoing an erroneous transition) and possibly shifting back at a later time to orbiting the same fixed point on the reference orbit. The lower the secondary minimum, the smaller the number of oscillations missed by the corresponding orbit. For instance, the three minima A, B, C of Fig. 4b correspond to orbits that miss only one oscillation of the actual orbit. As for the position of the secondary minimum along the unstable direction, the closer it is to the absolute minimum, the later the first "miss" occurs in the corresponding orbit.

More precisely, following Sparrow (1982), the sequence of oscillations of the cost function around the two fixed points of the system can be described by an infinite string of binary characters of the form *XXYXYY*..., where each symbol represents one oscillation around either one of the fixed points. Each secondary minimum, together with its basin of attraction, corresponds to orbits which are all described by the same string of characters over the period τ. Each time the assimilation period includes one more oscillation (the period of an oscillation varies itself with time), thereby opening one more possibility for one additional miss, the number of secondary minima is multiplied by two. All orbits originating from the attraction basin of the absolute minimum, and those orbits only, possess the same character string, over the entire assimilation period, as the reference orbit.

In a practical situation, it will be very important to ensure that the starting point of the minimization of the cost-function is located within the attractive basin of the absolute minimum, which guarantees that the string is correctly reproduced. One possible procedure for doing so is to increase stepwise the length of the assimilation period, to determine at each step the orbit minimizing the corresponding cost-function, and to use that orbit as starting point for the minimization at the next step. These quasi-static adjustments of the solution

thus help to control the assimilation process in the case of long assimilation periods. This method is developed in Section 4.

Coming back now to Fig. 4b, the dashed curve shows the variations of the error-contaminated forward cost-function J' (see eq. 2.3–4). The "observational noise" used to obtain J' was white in time and isotropic in space, with $\sigma^2 = 25$ along each direction, corresponding to a total variance $E^2 = 75$. It is seen that J' differs from the error-free cost-function J'_e by approximately that amount, and that the positions of the minima along the unstable direction are the same for both cost-functions. This is in agreement with what was anticipated from eq. (2.4), and strongly suggests that the position in phase space of the absolute minimum of J' must be a good approximation of the actual state of the system.

This "binary" description of the fluctuations of the cost-function is somewhat tentative, but corresponds to what has also been observed with the two-dimensional Hénon mapping (not shown). It clearly has limitations. It can be valid only if the sampling period δt is small enough to resolve each oscillation of the actual orbit. Also, it can apply only to orbits which oscillate in phase with the reference orbit, so that transition times occur simultaneously for the reference and perturbed orbits. In this sense, it can be valid only in some neighborhood of the absolute minimum, and certainly not over the whole phase space.

3. The error on the assimilated states

3.1. *The tangent linear assumption and the estimation error covariance matrix*

We have anticipated in the previous section that it should be possible, when using progressively older observations, and through a quasi-static control of the minimization procedure, to keep solutions in the basin of the absolute minimum of the cost function. We have also shown that the trajectories associated with such solutions remain close to the actual trajectory over the whole assimilation period. Provided that type of control is possible, the evolution of associated errors, $\delta x(t) = \hat{x}(t) - x(t)$ is approximately governed by the tangent linear equations resulting from the linearisation of eq. (2.1):

$$\frac{\mathrm{d}(\delta x(t))}{\mathrm{d}t} = \frac{\partial F}{\partial x}(x(t))\,\delta x(t). \tag{3.1}$$

This system is linear with variable coefficients, so that the transformation of any initial vector $\delta x(0)$ into $\delta x(t)$ by the tangent flow is given by a linear operator $H(t, x)$, called the *resolvent* of system (3.1) between times 0 and t:

$$\delta x(t) = H(t, x)\,\delta x(0). \tag{3.2}$$

Using these notations, the backward cost function (2.4) can be rewritten as a function of $\delta x = \delta x(0)$,

$$J(\tau, x, \delta x) = \frac{1}{N+1}\sum_{i=0}^{N} \|\delta x(t_i) - \varepsilon(t_i)\|^2. \tag{3.3}$$

Decomposing the right-hand side as in eq. (2.4) yields

$$\begin{aligned} J(\tau, x, \delta x) = {} & \frac{1}{N+1}\langle \delta x,\, U(\tau, x)\,\delta x\rangle \\ & + \frac{1}{N+1}\sum_{i=0}^{N} \|\varepsilon(t_i)\|^2 \\ & - \frac{2}{N+1}\langle V(\tau, x), \delta x\rangle, \end{aligned} \tag{3.4}$$

where the matrix $U(\tau, x)$ is defined as

$$U(\tau, x) = \sum_{i=0}^{N} H^{\mathrm{T}}(t_i, x)\,H(t_i, x), \tag{3.5}$$

and the vector $V(\tau, x)$ is defined as

$$V(\tau, x) = \sum_{i=0}^{N} H^{\mathrm{T}}(t_i, x)\,\varepsilon(t_i). \tag{3.6}$$

In the linear tangent approach, the cost function defined in eq. (3.3) has a unique minimum δx^*. This minimum can be obtained by differentiating eq. (3.4) with respect to δx. After a few lines of algebra, one obtains

$$\delta x^* = U(\tau, x)^{-1}\, v(\tau, x). \tag{3.7}$$

Considering that the observation error is a random process, the covariance matrix of the assimilation error is $C(\tau, x) = E(\delta x^*\, \delta x^{*\mathrm{T}})$, $E(\ )$ being the expectation operator. After substitution of eqs. (3.6) and (3.7) into the expression for the

covariance matrix, one easily obtains, using the assumption that the observation error $\varepsilon(t)$ is uncorrelated in time,

$$C(\tau, x) = U(\tau, x)^{-1} \times \left(\sum_{i=0}^{N} H^{\mathrm{T}}(t_i, x)\, \Sigma H(t_i, x) \right) U(\tau, x)^{-1}, \quad (3.8)$$

where Σ is the covariance matrix of the observation error $\varepsilon(t)$. Under the above assumption of isotropy of the error, $C(\tau, x)$ reduces to

$$C(\tau, x) = \sigma^2 U(\tau, x)^{-1}. \quad (3.9)$$

3.2. *The case of a tangent linear model with constant coefficients*

Let us consider the simple example of a tangent linear system with constant coefficients, possessing one stable and one unstable directions. Denoting by δx and δy the components of the state vector perturbation along the unstable and stable directions respectively, the system equations read:

$$\mathrm{d}\,\delta x/\mathrm{d}t = a\,\delta x, \quad (3.10a)$$

$$\mathrm{d}\,\delta y/\mathrm{d}t = -b\,\delta y, \quad (3.10b)$$

where a and b are positive constants. The matrix $U(\tau, x)$ is diagonal, which means that the estimation errors in the directions δx and δy are uncorrelated, their variances being equal to:

$$E(\delta x^{*2}) = \sigma^2 \frac{1 - e^{-2a\,\delta t}}{1 - e^{-2a(\tau + \delta t)}}, \quad (3.11a)$$

$$E(\delta y^{*2}) = \sigma^2 \frac{1 - e^{2b\,\delta t}}{1 - e^{2b(\tau + \delta t)}}. \quad (3.11b)$$

Simple conclusions can be drawn from these expressions. First of all, both $E(\delta x^{*2})$ and $E(\delta y^{*2})$ decrease as τ increases, which means that the inclusion of remote observations in the assimilation does improve the quality of the estimation. The asymptotic values of the estimation error variances are:

$$\lim_{\tau \to \infty} E(\delta x^{*2}) = \sigma^2 (1 - e^{-2a\,\delta t}), \quad (3.12a)$$

$$\lim_{\tau \to \infty} E(\delta y^{*2}) = 0. \quad (3.12b)$$

As τ goes to infinity, the error variance along the unstable manifold does not converge to zero, a situation that is not favorable for future prediction. Yet the error is reduced relative to the case where no past data are assimilated. The reduction factor is determined by the density of observations. For $a\,\delta t \ll 1$ this factor is roughly proportional to the sampling rate of the observations δt. We now prove that the same properties hold in the general case.

3.3. *The case of a general tangent linear model*

The estimation error covariance $C(\tau, x)$ is a symmetric, positive-definite matrix for finite values of τ. The inner product $\langle \delta x, C(\tau, x)\, \delta x \rangle$, where $\|\delta x\| = 1$, is the variance of the assimilation error along the direction δx. In the Appendix, we prove three important properties; the first two are rigourously true under general assumption, but the third one can only be proven for uniformly hyperbolic systems (Guckenheimer and Holmes, 1983). These properties concern the behaviour of the covariance matrix as τ goes to infinity:

(i) The function $\langle \delta x, C(\tau, x)\, \delta x \rangle$ is a monotonically decreasing function of τ, for any given direction δx in the tangent space. This property implies that, as in the constant-coefficient case, the variance of the assimilation error always decreases as the assimilation period is extended backward in time.

(ii) The error covariance matrix $C(\tau, x)$ has a limit matrix $C_\infty(x)$ as $\tau \to \infty$.

(iii) $C_\infty(x)$ is a singular non-zero matrix, meaning that only some components of the assimilation error tend to vanish as one incorporates more and more remote past observations in the assimilation process. The kernel of $C_\infty(x)$ is the subspace orthogonal to the unstable manifold of the tangent linear system.

The last property confirms the conclusions drawn from the phenomenological study of Section 2. As τ goes to infinity, the component of the error orthogonal to the unstable manifold vanishes, while the component along that manifold does not, which implies that the assimilated solution does not as a general rule converge to the exact solution.

These results can be transposed to assimilation of future observations, for which the projection of the error onto the subspace orthogonal to the stable manifold goes to zero. In order to reduce the

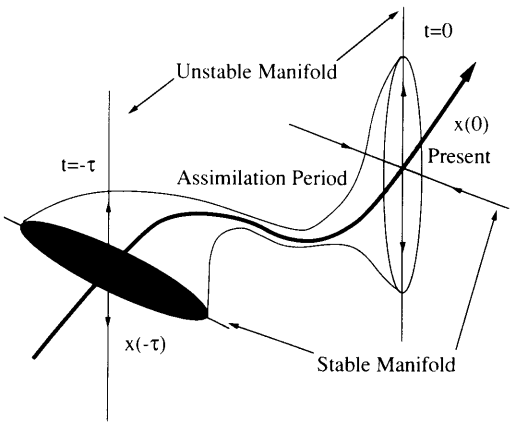

Fig. 6. Schematic representation, in phase space, and along the assimilation period, of the dispersion of trajectories associated to the absolute minimum of the cost function. The heavy solid curve represents the exact trajectory. The stable and unstable manifolds are shown at the beginning and at the end of the assimilation period (note that in general these manifolds are not orthogonal). The ellipses represent the dispersion of the error associated to the minimizing trajectories. This error spreads along the stable manifold at the beginning of the assimilation period and along the unstable manifold at the end. In the middle of the period, the error is small in all directions.

error in all directions and to perform an "optimal shadowing" of the system's trajectory at $t=0$, one has to perform variational assimilation with both past and future observation (Farmer and Sidorovich, 1991). The schematic diagram of Fig. 6 illustrates the behaviour of the estimation error along the assimilation period. The assimilation error concentrates along the stable manifold at the beginning of the assimilation period, and along the unstable manifold at the end of the period.

3.4. Upper bounds for the assimilation error

The direction in which the assimilation error δx^* has the largest variance is the eigendirection associated to the largest eigenvalue of $C(\tau, x)$. This eigenvalue is by definition the *spectral radius* of $C(\tau, x)$ and will be denoted $\mu(C(\tau, x))$. We now attempt to determine bounds for that quantity. We shall denote $\mu(A)$ and $\nu(A)$ the largest and smallest eigenvalues of a symmetric non-negative matrix A. From eq. (3.9), $\mu(C(\tau, x)) = \sigma^2/\nu(U(\tau, x))$. It is easy to deduce from (3.5) the following inequality:

$$\mu(C(\tau, x)) \leqslant \sigma^2 \left(\sum_{i=0}^{N} \nu(H^{\mathrm{T}}(t_i, x)\, H(t_i, x)) \right)^{-1} \tag{3.13}$$

The term with index i in the sum is the smallest amplification factor (usually less than 1), between times 0 and t_i, of a perturbation governed by the tangent linear model. Hence, it is equal to the inverse of the growth rate of the fastest growing mode (see, e.g., Lorenz, 1965; Lacarra and Talagrand; 1988, Molteni and Palmer, 1993) between the negative time t_i and 0. Since it is a positive number, it can be put into exponential form,

$$\nu(H^{\mathrm{T}}(t_i, x)\, H(t_i, x)) = \exp(2\lambda_i(x)\, t_i), \tag{3.14}$$

and the inequality (3.13) can be written

$$\mu(C(\tau, x)) \leqslant \sigma^2 \left(\sum_{i=0}^{N} e^{2\lambda_i(x)\, t_i} \right)^{-1} \tag{3.15}$$

The "local" exponents $\lambda_i(x)$ are positive and tend, as i increases (independently of x) to the largest positive Lyapunov exponent. This quantity can be obtained by the asymptotic singular values of the resolvent matrix (Ruelle, 1989). Thus, the inclusion of remote past data has an exponentially decreasing impact on the reduction of the error. The terms contributing most to this reduction are the first terms of sum in the r.h.s. of (3.15) (the first term is 1).

Since the exponents $\lambda_i(x)$ are local measures of the instability, the reduction should be largest just after the system error has undergone a period of weak instability, in which case the first few exponents are small. In the opposite case, when the system has undergone a period of strong instability, the reduction is inefficient, since the first exponents in the series (3.15) are large. Another conclusion that can be drawn from inequality (3.15) is that it is important to have a high density of observations, resulting in many terms in the sum.

The tail of the sum in (3.15) becomes negligibly small when, say, $\exp(2\lambda_i t_i) \leqslant 0.01$. Thus, it is worthless to take an assimilation period longer than the *efficient assimilation period*, $\tau_{\mathrm{ef}}(x) = -\mathrm{Ln}(0.01)/(2\lambda(x))$, where $\lambda(x)$ is a typical value of the local exponents. The efficient period, which may vary with time, can be estimated numerically in the course of the assimilation process, by

calculating the largest local growth rate, as was done experimentally in order to determine the predictability of particular weather situations (Molteni and Palmer, 1993). A somewhat cruder estimate can be obtained by replacing $\lambda_i(x)$ by the largest Lyapunov exponent of the system, Λ_1, leading to $\tau_{ef} = -\mathrm{Ln}(0.01)/(2\Lambda_1)$. For the Lorenz system, $\Lambda_1 \approx 1$ and $\tau_{ef} \approx 2$. Numerical results will confirm this estimate in the next Sections. If one assumes for the large-scale atmospheric flow an average doubling time of small errors of 2.5 days, as it was diagnosed by Lorenz (1969), the efficient assimilation period for large-scale flow should be of the order of a week. However, this estimation must not be taken too strictly. Many of the assumptions made here, like, for instance, the assumption of a perfect model, will not be verified due to the quality of the atmospheric models in a more realistic situation.

4. Quasi-static variational assimilation with the Lorenz system

The goal of this section is two-fold: On the one hand, it is designed to verify numerically the statements of the previous section, by performing assimilation experiments with the Lorenz system. On the other hand, we develop here the quasi-static variational assimilation (QSVA) algorithm proposed in Section 2, which allows to track the absolute minimum of the cost function over long assimilation periods. Other interesting efficient algorithms, such as the *simulated annealing* (Robertson et al., 1989), call for an application to this highly nonlinear variational problem where the absolute minimum is sought. However, as we shall see below, QSVA allows a natural control of the absolute minimum. By contrast, in the simulated annealing approach, only prescribed control parameters decide whether the algorithm is to be stopped, in which case the solution can fall off far away from the actual global minimum, even though the value of the cost function is very low.

The Lorenz system is integrated, as above, with a predictor-corrector temporal scheme, and a time step of 0.02. Minimization is achieved by using a Quasi-Newton algorithm (see Vautard and Legras (1988) for technical details). The gradient of the cost function is calculated using the adjoint of the model (Talagrand and Courtier, 1987). The adjoint equations were already used for the Lorenz system by Gauthier (1992) and Stensrud and Bao (1992). The reference orbit to be reconstructed corresponds to the time sequence presented in Fig. 2. The "present" is time $t = 0$. As above, the observation error is modeled by an isotropic Gaussian white noise process with unit covariance matrix in the coordinates x, y, z ($E^2 = 3$). All experiments described hereafter are performed with an observation sampling rate $\delta t = 0.1$.

4.1. *Quasi-static variational assimilation*

For numerical reasons which have already been explained, the minimization is performed with respect to the state $\hat{x}(-\tau)$ at the initial the time of the assimilation period. The QSVA algorithm can be decomposed into three steps:

(1) Initialisation: Use the observation $\hat{x}^{(0)} = \tilde{x}(-\tau)$ as starting point for the minimization of the forward cost function $J'(\delta\tau, \hat{x}(-\tau), x(-\tau))$, defined over the time interval $(-\tau, -\tau + \delta\tau)$. The minimizer thus obtained is denoted $\hat{x}^{(1)}$.

(2) From stage i to stage $i+1$: Use $\hat{x}^{(i)}$ as starting point for the minimization of the cost function $J'((i+1)\,\delta\tau, \hat{x}(-\tau), x(-\tau))$, defined over the interval $(-\tau, -\tau + (i+1)\,\delta\tau)$. This leads to the minimizer $\hat{x}^{(i+1)}$. Index i varies from 1 to $M-1$, hence $\tau = M\,\delta\tau$.

(3) Final stage: the solution $\hat{x}^{(M)}$ of the last minimization is intended to be the absolute minimizer of $J'(\tau, \hat{x}(-\tau), x(-\tau)) = J(\tau, \hat{x}(0), x(0))$. In order to obtain the corresponding estimated state vector $\hat{x}(0)$ at time 0, the model must be integrated over the assimilation period, with $\hat{x}^{(M)}$ as initial condition at time $-\tau$. Note that $M = N\,\delta t/\delta\tau$.

In the experiments to be described now, the increment is taken as $\delta\tau = 0.2$. QSVA is performed with 100 independent realisations of the observation error, for $\tau = 0$, 1, 3, and 8. The value $\tau = 0$ corresponds to a case where no assimilation is in effect performed, the "assimilation" error being thus identical to the observation error.

In Figs. 7a–d we display the cloud of the 100 assimilated solutions at time 0, for the various values of τ. As in Fig. 3, the position in phase space is defined by a projection onto the subspace spanned by the (linear) stable and unstable manifolds (note that the scales are different, in both directions, in panel (a) and in panels (b–d)). As the assimilation period τ increases, the cloud of

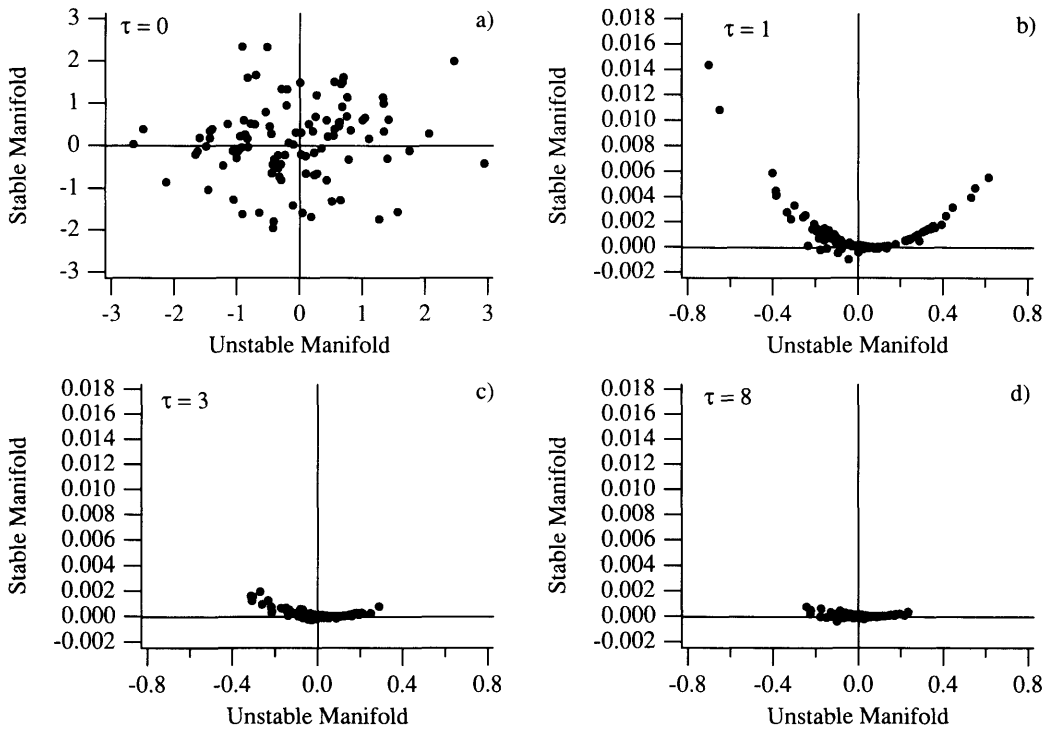

Fig. 7. Projection of the 100 minimizing solutions, at the end of the assimilation period, onto the plane spanned by the stable and unstable directions, defined as in Fig. 3. Values of τ are indicated on the panels. The projection is not an orthogonal projection, but a projection parallel to the local velocity vector $(dx/dt, dy/dt, dz/dt)$ (central manifold).

points shrinks in both directions, and tends to concentrate, as expected from the theoretical arguments of Sections 2 and 3, along the unstable direction. Along that direction, the diameter of the cloud is reduced by a factor of about 10 between $\tau = 0$ and $\tau = 3$. However, in agreement with the estimate $\tau_{ef} \approx 2$ of the previous section, no really significant change occurs when τ increases beyond 3. It can also be noted that the local curvature of the unstable manifold is clear on panel (b).

Tables 1a–c show the eigenvalues and eigenvectors of the assimilation error covariance matrix, for $\tau = 3$, as estimated directly from the numerical sample of 100 points (Table 1a), from the theoretical formulae (3.5) and (3.9) (Table 1b) and from the assimilated solutions obtained from a direct variational assimilation starting from the raw observations without the incremental process (Table 1c). The good agreement between the first two sets of estimates shows that QSVA is capable of maintaining the system within the tangent linear regime. By contrast, raw assimilation provides a totally erroneous estimation of the covariance matrix. Note that the largest eigenvalue in this case is three orders of magnitude larger than in the other cases. This is due to the fact that, at the end of the assimilation period, the 100 trajectories have progressively spread all over the attractor.

Table 2 shows various estimates of the spectral radius $\mu(C(\tau, x))$ for different values of τ. The agreement is very good between the values obtained from the observed dispersion of the QSV-assimilated states and from the tangent linear equations, but the upper bound (3.13) grossly overestimates the spectral radius for long assimilation periods. Between $\tau = 0$ and $\tau = 8$, the square error is reduced by a factor of about 70, which is much more than the value ($\simeq 5$) obtained from the theoretical formula (3.15). The large difference between the upper bound and the actual estimates is due to the eigenvalue majoration of the matricial

Table 1. *Eigenvectors* (w_1, w_2, w_3) *and associated eigenvalues of the covariance matrix of the estimation error at the final time of the assimilation period, and for* $\tau = 3$

(a) the covariance matrix is estimated from the ensemble of 100 assimilations performed with the QSVA algorithm.

	x-component	y-component	z-component	Eigenvalue
w_1	0.397	0.462	0.793	3.3×10^{-2}
w_2	0.318	0.741	−0.591	5.2×10^{-3}
w_3	0.861	−0.486	−0.147	6.1×10^{-8}

(b) the covariance matrix is estimated through the tangent linear system and eqs. (3.5)–(3.9).

	x-component	y-component	z-component	Eigenvalue
w_1	0.400	0.471	0.786	2.91×10^{-2}
w_2	0.311	0.736	−0.600	6.71×10^{-3}
w_3	0.861	−0.486	−0.147	3.5×10^{-12}

(c) the covariance matrix is estimated from the 100 assimilation experiments performed without the QSVA algorithm, i.e., the minimization is performed directly over the whole assimilation period using the raw observations at time $-\tau$ as the starting point of the minimization.

	x-component	y-component	z-component	Eigenvalue
w_1	0.550	0.831	0.008	29.4
w_2	0.009	0.003	−0.995	12.1
w_3	0.829	−0.555	−0.060	0.181

sum forming U in eq. (3.15). The assimilation error increases drastically once the obtained states fall off the right "valley" of the cost-function. This happens for $\tau \geqslant 2$. For $\tau = 1$, the cost function is nearly quadratic and most solutions remain near the exact solution. For $\tau = 8$, the cloud of points is spread all over the attractor (at the end of the assimilation period).

In order to study the time variations of the reduction factor along the unstable manifold, we calculate the theoretical value of $\mu(C(\tau, x(t))$ (as given by formula (3.9)), for a time t varying from 0 to 22, and for an assimilation period τ varying from 0 to 5. At a given time t, the reduction factor $\mu(C(\tau, x(t)))$ is a monotonically decreasing function of τ. In Fig. 8a we display, as functions of t, the values of τ at which a reduction factor $\mu = 0.8$, 0.5, 0.2, 0.1, and 0.05 is reached. In Fig. 8b, the value of the first exponent $\lambda_1(x(t))$ in the r.h.s. of (3.15) is plotted as a function of time t. This first exponent is a measure of the recent-past instability. Several conclusions can be drawn from these figures. First, the peaks of both graphs coincide, meaning that, as expected, the assimilation error is not much reduced during or at the end of a period of strong instability. By contrast, the error is greatly reduced at times t when $\lambda_1(x(t))$ is small. Strong instantaneous instability typically occurs, in the Lorenz system, when the trajectory lies close to the origin, which typically occurs near $t = 12.1$, or $t = 17.1$, one cycle before the transitions observed at $t = 12.8$ and $t = 17.8$. Therefore, during and immediately after an instability period, one should not expect assimilation to produce accurate estimated states, and the use of long assimilation periods does not help. At later times, when the flow

Table 2. *Estimates of the spectral radius* $\mu(C(\tau, x(t)))$ *of the estimation error covariance matrix at the end of the assimilation period, and for* $\tau = 0$, 1, 2, 3 *and* 8

$\mu(C(\tau, x))$	Cloud of points QSVA	Cloud of points raw assimilation	Linear tangent system	Upper bound
$\tau = 0$	1	1	1	1
$\tau = 1$	0.36	0.37	0.39	0.46
$\tau = 2$	5.9×10^{-2}	5.74	4.5×10^{-2}	0.401
$\tau = 3$	3.3×10^{-2}	29.4	2.9×10^{-2}	0.397
$\tau = 8$	1.4×10^{-2}	59.9	*	0.396

In the left column, the estimates are calculated from the ensemble of 100 assimilations (see also Fig. 7). The 2nd column contains the values obtained from the raw assimilation. In the 3rd column, the estimates are obtained from the tangent linear system and eqs. (3.5–3.9) (the star indicates a computational overflow). The estimates in the right-hand column are the upper bounds defined by eq. (3.13).

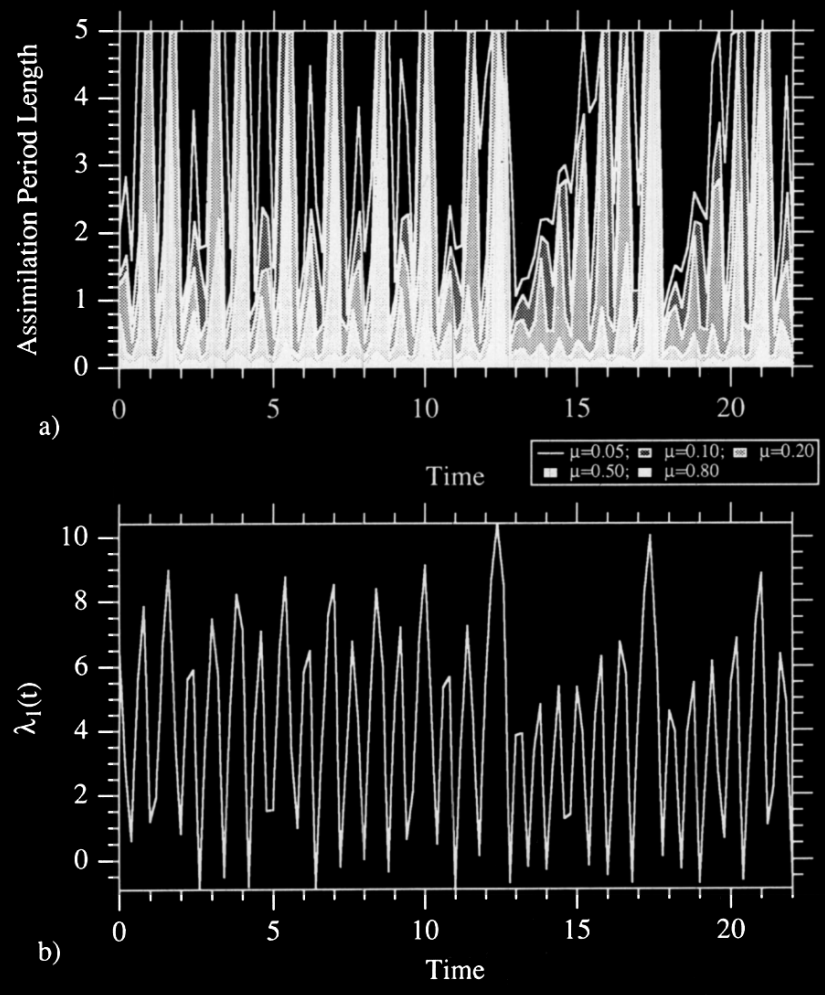

Fig. 8. (a) Values, as functions of time t, of the length τ of the assimilation period for which the spectral radius $\mu(C(\tau, x(t)))$ of the estimation error covariance matrix reaches the values $\mu = 0.8, 0.5, 0.2, 0.1, 0.05$, (from darker to lighter shading). The spectral radius is calculated from the linear tangent approximation of the covariance matrix (eqs. (3.5)–(3.9)). (b) For the same time period, values of the first exponent $\lambda_1(x(t))$ in the r.h.s. of (3.15).

becomes more stable, the error can be drastically reduced by using QSVA.

It must finally be stressed that there are limitations, both numerical and observational, to QSVA. The control of the absolute minimum can be maintained in the course of the quasi-static adjustments only if the starting point of each new minimization is located within the basin of attraction of the absolute minimum. This requires, at least, that both the observational error and the sampling interval of observations be small enough. In the case of the Lorenz system, the observations must at least be sufficiently accurate and frequent in time to resolve each individual oscillation around either one of the fixed points.

4.2. Sequential quasi-static variational assimilation

The algorithm described above has a major drawback, namely its requirement for repeated minimizations over overlapping periods of increasing length associated with a high computational cost. One way to reduce that cost in a practical situation would be to proceed as follows. Assume the absolute minimizing solution of the

forward cost-function $J'(\tau, \hat{x}, x(t-\tau))$ has been determined. This cost function corresponds to the interval $(t-\tau, t)$, where t is the present time. This minimizing solution is denoted $\hat{x}(t-\tau)$. New observations are acquired over the time interval $(t, t+\delta\tau)$, where $\delta\tau < \tau$. The new cost-function $J'(\tau, \hat{x}, x(t-\tau+\delta\tau))$ is then minimized, with $\hat{x}(t-\tau+\delta\tau)$ as starting point of of the minimization. This point is obtained from $\hat{x}(t-\tau)$ by integrating (2.1) for a time $\delta\tau$. The new minimizing solution defines an initial state for a forecast starting at time $t+\delta\tau$. The process can be repeated, each new minimization being performed over a time interval shifted by $\delta\tau$ from the time interval of the previous minimization. This will work under the condition that, at every step, the computed minimum is the absolute minimum, and lies within the attractive basin of the absolute minimum of the next minimization.

A large variety of schemes for *sequential quasi-static variational assimilation* of this type can be defined, only by varying the two parameters τ and $\delta\tau$. A number of meteorological centers have decided to develop variational assimilation for operational purposes. The schemes which are being developed can be described as forms of sequential quasi-static variational assimilation. In these schemes as they are defined at present (see, e.g., Courtier et al., 1994), there is no overlap between successive assimilation periods ($\delta\tau = \tau$), but the solution obtained at each step is used not only as the starting point for the next minimization, but is actually included, as a "background" term, into the new cost-function to be minimized. This guarantees that, at each step, the information contained in the observations performed before the present assimilation period is not lost.

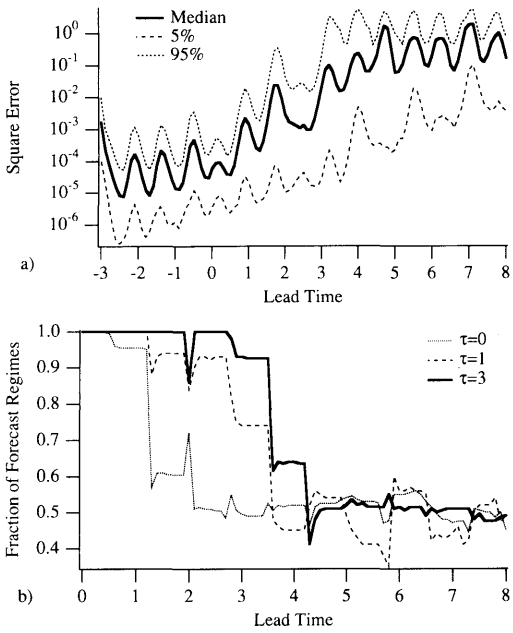

Fig. 9. (a) Median (heavy curve), and 5- and 95-percentiles of the distribution of the logarithms of the normalized squared assimilation and forecast errors, as obtained from 100 independent realisations of the observational noise, and as functions of forecast time. The assimilation period $(-3, 0)$ is included on the time axis. (b) Fraction of successful regime forecasts as a function of the forecast time, for $\tau = 0$ (no assimilation), 1 and 3.

5. Variational assimilation and predictability

5.1. *Instantaneous predictability*

A statistical measure of the quality of an assimilation process can be obtained from the distribution, for an ensemble of realizations of the observational noise, of the error in the forecasts produced from the assimilated states. Fig. 9a, obtained from the QSVA experiments described in the previous Section, shows the 5-, 50- and 95-percentiles of the distribution of the logarithms of the squared forecast error, calculated from the sample of 100 realizations of the observational noise already used for producing Fig. 7. The length of the assimilation period is $\tau = 3$. Note that the assimilation period is included in the figure, and that only positive values along the horizontal time axis correspond to "forecasts". The square error on the vertical axis is normalized by twice the total variance of the vector $(x(t), y(t), z(t))$, so that the averaged squared distance between two points randomly chosen on the attractor is equal to 1. All predictability is lost when the distance between the forecast and the actual values reaches that value. The oscillations of the three curves in the figure are in phase, which means that they are statistically significant. The least successful forecasts, associated with the 95-percentile, reach the predictability limit for a lead time of about 3 time units. The median value of the forecast error (50-percentile) reaches the predictability limit for a lead time of about 4 to 5 time units. This range can

be taken as the average predictability time for the particular point on the attractor considered here, and for the particular choices made here as concerns the observational parameters.

As already mentioned, the sign of the variable $x(t)$ of the Lorenz system is an indicator of which of the two regimes the state of the system belongs to. Fig. 9b shows the fraction of successful regime forecasts as a function of the lead time, for the three values $\tau = 0$, 1 and 3 of the length of the assimilation period. All three curves saturate at the climatological value of 50% at lead time $t \approx 4.5$, in agreement with the estimate already obtained from Fig. 9a. When no assimilation is performed ($\tau = 0$), almost half the forecasts miss the transition at lead time 1.2 (see Fig. 2). For $\tau = 1$, a significant fraction of the forecasts undergo an erroneous transition at lead time $t = 2.8$. For $\tau = 3$, more than 90% of the forecasts are successful up to the transition occurring at time 3.6. The cusp present at time 2 is due to the fact that $x(t)$ is close to zero at that time, leading to difficulties in determining the regime. The differences between the various curves in Fig. 9b clearly show the importance, for the quality of subsequent forecasts, of performing assimilation over long periods, for which QSVA is required.

5.2. *Variability in predictability*

Palmer (1993) has shown that the predictability of the Lorenz system is subject to large variability. This aspect is now studied from a sequential QSVA experiment performed along the period (0,22), with again 100 realisations of the observation error, using $\delta t = 0.1$ and $\delta\tau = 0.2$. We define at each instant the predictability time, as the average, over the 100 realisations, of the time during which the sign of the forecast variable $x(t)$ remains the same as the actual one. The variations of the predictability time is displayed in Fig. 10 for three values of the assimilation period τ. It is larger right after the instability periods, precisely when long assimilation periods are not efficient (near $t = 12.5$ and $t = 17.5$, compare with Figs. 8a, b and with Fig. 2). By contrast, long assimilation periods ($\tau = 3$) and QSVA increase significantly the predictability time with respect to short assimilation periods ($\tau = 1$) when instability periods are relatively far in the past (here, more than 1 unit, say), and before new instability periods like at times 12 and 17). In such cases, the number of correctly reproduced transitions is significantly increased by increasing the assimilation period length, leading to a greater average predictability time. This shows clearly the importance of taking long assimilation periods in order to correctly forecast transitions. In general, after a long period of stability, the terms in the r.h.s. of eq. (3.15) decrease slowly, leading to a much longer efficient assimilation period τ_{ef}.

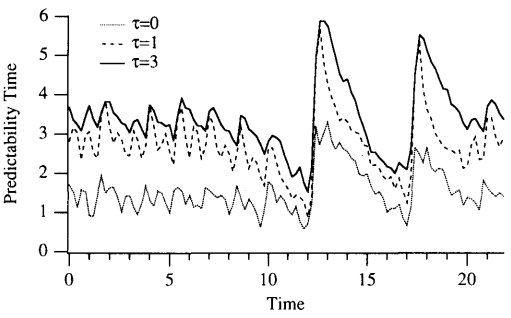

Fig. 10. Predictability time (see text for definition), as a function of the absolute time t, and for three values of the length τ of the assimilation period.

5.3. *Average predictability*

We now assess the impact of variational assimilation on average predictability. Eqs. (2.8) are integrated for a period of 1000 time units. Observations are generated, as above, at a sampling rate of $\delta t = 0.1$, with a unit error covariance matrix. Sequential QSVA is applied as above, with $\tau = 0$, 1, and 3 and $\delta\tau = 0.2$. Fig. 11a, whose format is similar to that of Fig. 9a, shows the average logarithm of the squared forecast error over these cases. Over the 1000 time units covered, a total of 5000 forecast experiments are taken into account, on average (sampled every $\delta\tau$ units).

We first note that averaging over a large number of different situations filters out the oscillations present in Fig. 9a. For $\tau = 3$, the error growth is exponential, with a rate which can be verified to roughly correspond to the first Lyapunov exponent of the system ($\Lambda_1 \simeq 1$). The average predictability time (i.e., the lead time at which the error growth saturates) is about 4-5 time units, in agreement with the estimate already obtained from Fig. 9a for one particular point of the attractor. For smaller assimilation periods τ, the growth rate

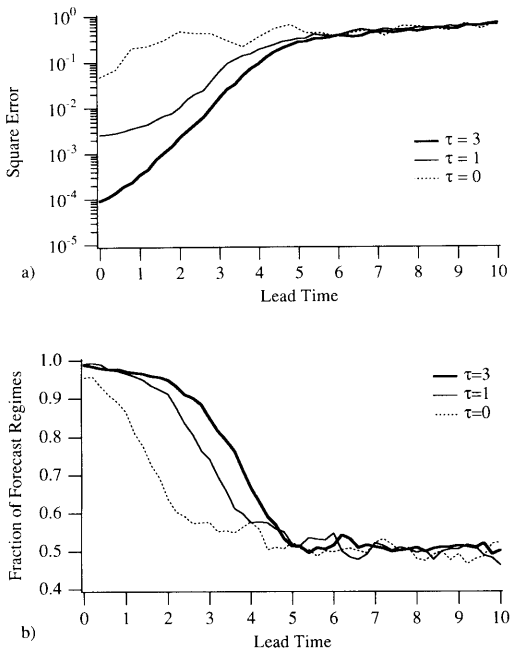

Fig. 11. (a) Squared normalized forecast error, averaged over the attractor, as a function of forecast time, and for the values $\tau = 0$, 1 and 3 of the length of the assimilation period (b). Fraction, over the attractor, of successful regime forecasts, as a function of forecast time, and for the values $\tau = 0$, 1 and 3 of the length of the assimilation period.

is smaller, even for the same value of the error (compare the $\tau = 3$ and the $\tau = 1$ curves). The difference between the growth rates is due to the fact that for large assimilation periods, the error is mostly concentrated along the unstable manifold, leading to a larger growth rate than when errors are somewhat spread in all directions. For $\tau = 1$, the predictability time decreases somewhat, and is equal to about 2 time units when no assimilation is performed ($\tau = 0$).

Fig. 11b, which is in the same format as Fig. 9b, shows the average fraction of successfully forecast regimes. All curves saturate at the climatological value (0.5) for a lead time of 5 units. There is a difference of about 0.7 time units in regime predictability between forecasts corresponding to $\tau = 1$ and to $\tau = 3$. At a lead time of 3, for instance, about 85% of the regimes are correctly forecast for $\tau = 3$ whereas the corresponding fractions are respectively 75% and 55% for $\tau = 1$ and for no assimilation.

6. Conclusion and discussion

We have performed a theoretical investigation on variational assimilation of noisy observations of a deterministic, nonlinear, chaotic dynamical system. The study is made under the hypothesis of a perfect model, and the results are illustrated by numerical experiments performed with the three-variable system introduced by Lorenz (1963). We have studied how the accuracy of the state obtained at the end of the assimilation period varies when the length of the period increases back to infinity. Mathematically, infinite accuracy is obtained on the component of the flow that projects along the stable manifold of the system, while finite accuracy is obtained on the component of the flow that projects along the unstable manifold. That latter component being the component that generates forecast errors, inclusion of past observations, even with a perfect model, can have only a limited impact on predictability.

The purpose of variational assimilation is to determine the trajectory of the system that best fits the observations over the assimilation period. This is achieved through minimization of an appropriate cost function. As the length of the assimilation period increases back to infinity, the chaotic nature of the system produces a foliation of the contours of the cost-function and generates infinitely many minima. The existence of multiple minima in the cost-function for the case of a simplistic atmospheric model was also shown by Li (1991) and Gauthier (1992). These minima are located in parallel narrow "valleys" elongated in the direction of the unstable manifold of the system at the final time of the assimilation period, the time at which forecasts are to be issued. The attractor of the system generally consists of a fractal foliation of infinitely many sheets parallel to the unstable manifold (this foliation is consequently infinitely more refined than the foliation of a cost-function defined over a finite time interval). The absolute minimum of the cost-function always lies near the sheet containing the exact state of the flow.

Starting the minimization from the observed values generally leads to a secondary minimum. We proposed here a quasi-static version of variational assimilation which, provided the temporal

density of observations is high enough, allows to track the absolute minimum over progressively longer assimilation periods. A sequential version of the quasi-static variational assimilation, suggested in Subsection 4.2, allows to reduce drastically the large computer time required for the quasi-static adjustments of the solution. It is to be noted that variational assimilation, as planned to be operationally implemented in several meteorological centers, can be described as one particular form of sequential quasi-static variational assimilation.

Comparison has been made of the quality of forecasts originating from initial conditions obtained through assimilation and from simply "observed" initial conditions. The conclusion is that assimilation significantly increases predictability. The longer the assimilation period, the higher the predictability. The gain in accuracy that can be obtained by increasing the length of the assimilation period is however limited, and strongly dependent on the instability of the flow over the recent past. We propose an approximative estimate for the length τ_{ef} of the "efficient" assimilation period. This length is proportional to the average doubling time of growth of small errors, found by Lorenz (1969) to be of the order of 2.5 days for the large-scale atmospheric flow. Using this value, a crude estimate of the efficient assimilation period is $\tau_{ef} \approx$ one week for the large-scale atmospheric flow.

However, the oversimplified assumptions made here (perfect model, for instance) cast doubt on the reliability of such an estimate. The use of an imperfect model will probably decrease the efficient assimilation period. The model errors have an increasing influence as assimilated observations get older. Nevertheless, it is generally agreed that the initial errors are the main reason of the predictability limitation of the forecasts in the short and medium terms, hence leaving us with the hope that assimilation periods of the order of several days could be considered. In any case, experiments with realistic models and actual observations have to be carried out in order to clarify this point.

Above all, if valid qualitative conclusions, based on arguments which are independent of the particular properties of the Lorenz system, can certainly be drawn for the atmosphere (as for instance, the existence of multiple minima in the cost-function, the concentration of the assimilation error in the unstable modes, or the variation of the quality of the assimilation with the current stability of the flow), further work is certainly necessary before reliable quantitative estimates can be obtained (for instance, on the maximum value δt permissible for the QSVA algorithm to work, and on the efficient assimilation period length).

Preliminary applications of the QSVA algorithm by the authors to a quasi-geostrophic model, have been performed, yielding efficient assimilation periods of the order of 5–10 days. In a recent work, Luong (personnal communication) also applied QSVA for the assimilation of altimetric data in a quasi-geostrophic oceanic model, and came out with efficient assimilation periods of about 100 days, a period too long for the application of direct variational assimilation.

Just as predictability itself, the accuracy of assimilated solutions varies significantly with the current state of the flow. We have shown that the instantaneous estimation error strongly depends on the degree of instability of the recent past. Situations of low instability near the end of the assimilation period are favorable to high accuracy in the ensuing assimilated state. On the opposite, unstable situations lead to a relatively large estimation error, and no much improvement is to be expected in such situations through the use of long assimilation periods. This would typically apply just after the atmosphere has undergone a transition of weather regime.

We have shown how the variance of the instantaneous estimation error can be numerically estimated, using the linear tangent model and its adjoint. This method is however expensive. Its cost could be reduced, following the ideas developed in Toth and Kalnay (1993, 1994), by restricting the computations to a few sequentially estimated *breeding* modes. These modes, which are among the unstable modes, must contain a large proportion of the estimation error.

We have considered in this paper only *variational* assimilation. But some of the conclusions obtained, such as the concentration of assimilation error in the unstable modes of the flow, or the fact that the assimilation error cannot be reduced to zero by increasing the assimilation period, are absolutely general. As regards the algorithms themselves, the only real question is whether variational assimilation, or variants of variational assimilation such as QSVA, are more efficient or

not than other algorithms, such as the sequential algorithms studied by Miller et al. (1994). This question will be addressed in future work.

Finally, one main conclusion of this paper is that assimilation of past observations essentially reduces the error along the stable components of the flow. A dual result is that assimilation of future observations would reduce the error along the unstable components. If observations are available both before and after the time at which the state of the flow is to be estimated, the assimilation error becomes small in all directions. This has two consequences. First, it emphasizes the importance, in a reanalysis process of past observations, of using observations performed both before and after the estimation time. Second, it shows that any available "future information" could be useful for assimilation. Such future information can be produced, for instance, by empirical models of the low-frequency components of the atmospheric system (Vautard et al., 1995, *Mon. Wea. Rev.*, submitted). Such models are able to predict characters of the flow that ordinary weather prediction models are apparently not able to predict, owing partly to their systematic errors in the long run. As such, the empirical models of the low-frequency components bring additional information that is contained neither in past observations nor in the assimilating model. Low-frequency oscillations, such as intraseasonal modes could be good candidates for such a process. This would mean that empirical forecasts could be considered as "future observations" and used just as any other observation in the real-time assimilation process. Therefore, a complete assimilation cycle would contain the empirical forecast itself.

7. Acknowledgments

We are thankful to Bernard Legras, W. Zeitlin, and Pierre Gauthier for the interesting and fruitful discussions we had. The permanence in France of Carlos Pires was supported by the grant BD/2023/92 of JNICT (Junta Nacional de Investigaçao Cientifica e tecnologica), Portugal. This work was partly sponsored by The Groupement de Recherche "Méthodes Variationnelles en Météorologie et Océanographie": CNRS/Meteo-France. The software for nonlinear optimization was provided by INRIA.

8. Appendix

The behaviour of the covariance matrix of the estimation error in the limit of an infinitely long assimilation period

We prove statements (i), (ii) and (iii) of Subsection 3.3. We will need lemmas, separated into several algebraic lemmas (AL) and one topological lemma (TL).

AL1: If A and B are symmetric positive-definite matrices, then the matrix $C = A^{-1} - (A+B)^{-1}$ is also symmetric positive-definite. The proof is easily obtained by starting from a Cholesky factorisation of A, $A = P^{\mathrm{T}}P$, where P is an invertible matrix, and noting that the eigenvalues of the symmetric matrix

$$I - (I + (P^{\mathrm{T}})^{-1} BP^{-1})^{-1}$$

are all positive.

AL2: If $A(t)$ is a one-parameter family of symmetric positive-definite matrices such that $A(t') - A(t)$ is positive for $t' < t$, then $A(t)$ has a limit A_∞ as t goes to infinity. This limit is positive, but may be singular. Then, for any value of the parameter t, $A(t) - A_\infty$ has positive, possibly zero, eigenvalues. The proof is obtained by taking the limit, for any vector u, of $\langle A(t)\,u, u\rangle$, which is a decreasing, positive function of the parameter t, and therefore converges to a non-negative number $a(u)$. The limit matrix coefficient $(A_\infty)_{ij}$ is obtained by taking the limit, as t goes to infinity, of

$$A_{ij}(t) = \tfrac{1}{4}(\langle A(t)(u_i + u_j), (u_i + u_j)\rangle - \langle A(t)(u_i - u_j), (u_i - u_j)\rangle) \qquad \text{(A1)}$$

where u_i and u_j are the basis vectors. The convergence of $A_{ij}(t)$ is guaranteed by the convergence of each term of the difference.

AL3: If A is a symmetric positive-definite matrix of order n, u and v are any vectors in R^n, then the following two inequalities hold:

$$\langle u, v\rangle^2 \leqslant \langle Au, u\rangle\langle A^{-1}v, v\rangle, \qquad \text{(A2a)}$$

$$\langle Au, v\rangle^2 \leqslant \langle Au, u\rangle\langle Av, v\rangle. \qquad \text{(A2b)}$$

These inequalities result from Schwarz's inequality applied with the scalar product defined by the matrix A.

AL4: If A is a symmetric matrix, of order n, its kernel $\mathrm{Ker}(A)$, and its image $\mathrm{Im}(A)$ are mutually

orthogonal and in direct sum: $\mathrm{Ker}(A) \oplus \mathrm{Im}(A) = \mathrm{R}^n$. This follows from the diagonalisation of symmetric matrices, $\mathrm{Ker}(A)$ being associated to zero eigenvalues and $\mathrm{Im}(A)$ to nonzero ones.

TL1: The Oseledec theorem (Oseledec, 1968). For a large class of dynamical systems in a compact manifold of dimension n, the quantity

$$\Lambda(\delta x) = \lim \left(\frac{1}{|t|} \ln(\|H(t, x)\, \delta x\|) \right), \qquad t \to -\infty \tag{A3}$$

exists and takes at most n values, $\Lambda_1 \geqslant \Lambda_2 \geqslant \cdots \geqslant \Lambda_n$, the so-called *Lyapunov exponents* of the system. $H(t, x)$ is the resolvent of the system. In the tangent space at point x, there exist subspaces $E_i(x) = \{\delta x: \Lambda(\delta x) \leqslant \Lambda_i\}$ such that $\mathrm{R}^n = E_1(x) \supseteq E_2(x) \supseteq \cdots \supseteq E_n(x)$. The *unstable manifold* of the linear tangent system is the subspace $E_r(x)$, where r is the smallest index such that $\Lambda_r < 0$. There exists an inequality similar to (2.6a), i.e., there are constants $R > 0$ and $0 < \eta < 1$, such that for any vector δx belonging to the unstable manifold,

$$\|H(t, x)\, \delta x\| \leqslant R\, \|\delta x\|\, \eta^{|t|}, \qquad \text{for} \quad t < 0. \tag{A4}$$

The real number η can be chosen as any positive number greater than $\exp(\Lambda_r)$. When the system is hyperbolic, the theorem of Hirsch and Pugh (1970), from which inequalities (2.6a, b) result, also shows that the unstable manifold of the tangent system, $E_r(x)$, is locally tangent to, and has the same dimension as, the unstable manifold $W_u(x)$ of the nonlinear system.

Proof of statement (i) of Subsection 3.3

The fact that $\langle C(\tau, x)\, \delta x, \delta x \rangle$ is a monotonically decreasing function of τ results from Lemma AL1: If $\tau = N\, \delta t \leqslant \tau' = N'\, \delta t$, then $N \leqslant N'$ and, from eq. (3.9) one obtains:

$$C(\tau, x) - C(\tau', x) = \sigma^2 \left\{ U(\tau, x)^{-1} - \left[U(\tau, x) + \sum_{i=N+1}^{N'} H^{\mathrm{T}}(t_i, x)\, H(t_i, x) \right]^{-1} \right\}; \tag{A5}$$

therefore $C(\tau, x) - C(\tau', x)$ is a symmetric positive matrix, which proves the statement.

Proof of statement (ii) of Subsection 3.3

Equality (A5) also shows that the one-parameter matrix family $C(\tau, x)$ (τ is the parameter) satisfies the hypothesis of Lemma AL2, which proves the statement.

Proof of statement (iii) of Subsection 3.3

We first show that $\mathrm{Ker}(C_\infty) \subseteq E_r^{\perp}(x)$ (the subspace orthogonal to the unstable manifold). Take any two vectors u in $\mathrm{Ker}(C_\infty)$, and v in $E_r(x)$. We first prove that $\langle u, v \rangle = 0$. From inequality (A2a) of Lemma AL3 and eq. (3.9), one obtains, for any value of τ:

$$\langle u, v \rangle^2 \leqslant \langle C(\tau, x)\, u, u \rangle \langle U(\tau, x)\, v, v \rangle. \tag{A6}$$

The definition of $U(\tau, x)$, for $\tau = N\, \delta t$, leads to

$$\langle U(\tau, x)\, v, v \rangle = \sum_{i=0}^{N} \|H(t_i, x)\, v\|^2, \tag{A7}$$

which is bounded as τ goes to infinity, from inequality (A4) of Lemma (TL1) and from the fact that $v \in E_r(x)$. The fact that $u \in \mathrm{Ker}(C_\infty)$ makes the term $\langle C(\tau, x)\, u, u \rangle$ go to zero as τ goes to infinity. Therefore, $\langle u, v \rangle = 0$.

Next, we prove that $\mathrm{Im}(C_\infty) \subseteq E_r(x)$. Let v be a nonzero vector in $\mathrm{Im}(C_\infty)$. By definition, there is a nonzero vector u such that $v = C_\infty u$. Then, inequality (A2b) of AL3 defines the following upper bound for the quantity

$$\langle v, U(\tau, x)\, v \rangle = \langle C_\infty u, U(\tau, x)\, C_\infty u \rangle:$$

$$\langle C_\infty u, U(\tau, x)\, C_\infty u \rangle^2 \leqslant \langle C_\infty u, u \rangle \langle C_\infty U(\tau, x)\, C_\infty u, U(\tau, x)\, C_\infty u \rangle. \tag{A8}$$

For any value of τ, $C(\tau, x) - C_\infty$ is a positive matrix (Lemma AL2). Therefore, the second term in the r.h.s. of (A8) is less than $\langle C(\tau, x)\, U(\tau, x)\, C_\infty u, U(\tau, x)\, C_\infty u \rangle$. From eq. (3.9), one obtains

$$\langle C_\infty u, U(\tau, x)\, C_\infty u \rangle^2 \leqslant \langle C_\infty u, u \rangle \langle C_\infty u, U(\tau, x)\, C_\infty u \rangle, \tag{A9}$$

and since $U(\tau, x)$ is positive definite, the last inequality can be divided by $\langle C_\infty u, U(\tau, x)\, C_\infty u \rangle$, which leads to, after reintroduction of v,

$$\langle v, U(\tau, x)\, v \rangle \leqslant \langle C_\infty u, u \rangle. \tag{A10}$$

The expansion (A7) of the l.h.s. of this equation shows that the corresponding series is bounded, and that $\|H(\tau, x)\, v\|$ vanishes as τ goes to infinity. This proves that v belongs to $E_r(x)$, and that $\mathrm{Im}(C_\infty) \subseteq E_r(x)$. From the two inclusions $\mathrm{Ker}(C_\infty) \subseteq E_r^{\perp}(x)$ and $\mathrm{Im}(C_\infty) \subseteq E_r(x)$, and from Lemma AL4, we finally obtain the proof of statement (iii).

REFERENCES

Bennett, A. F. 1992. *Inverse methods in physical oceanography*. Cambridge University Press, Cambridge, United Kingdom, 346 pp.

Bennett, A. F. and Budgell, W. P. 1989. The Kalman smoother for a linear quasi-geostrophic model of ocean circulation. *Dyn. Atmos. Oceans* **13**, 219–267.

Budgell, W. P. 1986. Nonlinear data assimilation for shallow water equations in branched channels. *J. Geophys. Res.* **91** (C9), 10,633–10,644.

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

Courtier, P., Derber, J., Errico, R., Louis, J.-F. and Vukicevic, T. 1993. Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus* **45A**, 341–357.

Courtier, P., Thépaut, J.-N. and Hollingsworth, A. 1994. A strategy for operational implementation of 4D-VAR, using an incremental approach. *Q. J. R. Meteorol. Soc.*, in press.

Daley, R. 1991. *Atmospheric data analysis*. Cambridge University Press, Cambridge, U.K., 460 pp.

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

Devaney, R., and Nitecki, Z. 1979. Shift automorphisms in the Hénon mapping. *Comm. Math. Phys.* **67**, 137–148.

Errico, R. M. and Vukicevic, T. 1992. Sensivity analysis using an adjoint of the PSU-NCAR mesoscale model. *Mon. Wea. Rev.* **120**, 1644–1660.

Evensen, G. 1992. Using the extended Kalman filter with a multilayer quasigeostrophic ocean model. *J. Geophys. Res.* **97** (C11), 17,905–17,924.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99** (C5), 10,143–10,162.

Farmer, J. D., and Sidorovich, J. J. 1991. Optimal shadowing and noise reduction. *Physica D* **47**, 373–392.

Gaspar, P. and Wunsch, C. 1989. Estimates from altimeter data of barotropic rossby waves in the northwestern atlantic ocean. *J. Phys. Oceanogr.* **19**, 1821–1844.

Gauthier, P. 1992. Chaos and quadri-dimensional data assimilation. A study based on the Lorenz model. *Tellus* **44A**, 2–17.

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

Ghil, M. and Malanotte-Rizzoli, P. 1991. Data assimilation in meteorology and oceanography. *Adv. Geophys.* **33**, 141–266.

Hénon, M. 1976. A two-dimensional mapping with a strange attractor. *Comm. Math. Phys.* **50**, 69–77.

Hirsch, M. W. and Pugh, C. C. 1970. Stable manifolds and hyperbolic sets. *Proc. Symp. Pure. Math.* **14**, 133–163.

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

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82D**, 35–45.

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

Li, Y. 1991. A Note on the uniqueness problem of variational adjustment approach to four-dimensional data assimilation. *J. of the Met. Soc. of Japan.* **69**, 581–585.

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

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

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

Lorenz, E. N. 1965. A study of the predictability of a 28-variable atmospheric model. *Tellus* **27**, 321–333.

Lorenz, E. N. 1969. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.* **26**, 636–646.

Miller, R. N., Ghil, M. and Gauthiez, F. 1994. Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.* **51**, 1037–1056.

Molteni, F. and Palmer, T. 1993. Predictability and finite-time instability of the northern winter circulation. *Quart. J. R. Meteor. Soc.* **119**, 269–298.

Nese, J. M., Dutton, J. A. and Wells, R. 1987. Calculated attractor dimensions for low-order spectral models. *J. Atmos. Sci.* **44**, 1950–1972.

Oseledec, V. I. 1968. A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.* **19**, 197–231.

Palmer, T. 1988. Medium and extended-range predictability and the stability of the Pacific/North American mode. *Quart. J. R. Meteor. Soc.* **114**, 691–713.

Palmer, T. 1993. Extended-range atmospheric prediction and the Lorenz model. *Bull. Am. Meteor. Soc.* **74**, 49–65.

Parrish, D. and Derber, D. 1992. The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.* **120**, 1747–1763.

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

Robertson, D. H., Brown, F. B. and Navon, I. M. 1989. Determination of the structure of mixed-xenon clusters using a finite-temperature, lattice-based Monte Carlo method. *J. Chem. Phys.* **90**, 3221–3229.

Ruelle, D. 1989. *Chaotic evolution and strange attractors.* Cambridge University Press, Cambridge, 95 pp.

Sheinbaum, J. and Anderson, D. L. T. 1990a. Variational assimilation of XBT data. Part I. *J. Phys. Oceanogr.* **20**, 672–688.

Sheinbaum, J. and Anderson, D. L. T. 1990b. Variational assimilation of XBT data. Part II. Sensitivity studies and use of smoothing constraints. *J. Phys. Oceanogr.* **20**, 689–704.

Sparrow, C. 1982. *The Lorenz equations.* Springer-Verlag, New-York, 269 pp.

Stensrud, D. J. and Bao, J. W. 1992. Behaviors of variational and nudging assimilation techniques with a chaotic low-order model. *Mon. Wea. Rev.* **120**, 3016–3028.

Talagrand, O. 1992. Data assimilation problems. In: Proceedings, NATO Advanced Study Institute, *Energy and water cycles in the climate system*, (Glücksburg, Germany, October 1991), E. Raschke and D. Jacob, eds. Forschungszentrum Geesthacht, Geesthacht, Germany, 187–213.

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.

Thacker, W. C. and R. B. Long, 1988. Fitting dynamics to data. *J. Geophys. Res.* **93**, 1227–1240.

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

Tibaldi, S. and Molteni, F. 1990. On the operational predictability of blocking. *Tellus* **42A**, 343–365.

Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NMC. The generation of perturbations. *Bull. Amer. Meteor. Soc.* **74**, 2317–2330.

Toth, Z. and Kalnay, E. 1994. *Ensemble forecasting at NMC. The use of the breeding method for generating perturbations.* Proceedings, 10th AMS Conference on Numerical Weather Prediction, Portland, Oregon, USA, July 1994. Published by the American Meteorological Society.

Vautard, R. and Legras, B. 1988. On the source of midlatitude low-frequency variability. Part II. Nonlinear equilibration of weather regimes. *J. Atmos. Sci.* **45**, 2845–2867.

Vautard, R. 1990. Multiple weather regimes over the north Atlantic: analysis of precursors and successors. *Mon. Wea. Rev.* **118**, 2056–2081.

Zou, X. Navon, I. M. and Le Dimet, F.-X. 1992. Incomplete observations and control of gravity waves in variational data assimilation. *Tellus* **44A**, 273–296.