Tellus (1998), 50A, 58–75
Printed in UK – all rights reserved

TELLUS
ISSN 0280–6495

Convergence of data assimilation by periodic updating in simple Hamiltonian and dissipative systems

By A. HANNACHI* and K. HAINES, *Department of Meteorology, Edinburgh University, Edingburgh EH9 3JZ, UK*

(Manuscript received 27 December 1996; in final form 28 August 1997)

ABSTRACT

In this paper, we study the influence of the interval between data insertion events on the convergence of sequential data assimilation problems. An example of a conservative Hamiltonian system is presented (that of Hénon and Heiles (1964)) where sequential assimilation with periodic data insertion every Δt achieves a more rapid convergence if data is not inserted at the smallest possible update interval, Δt. It is shown analytically that this is true for all Hamiltonian systems when the updated variables produce convergence of the assimilation, because the resolvent matrix then varies as $O(\Delta t^2)$ to highest order. The theory successfully predicts the turnover point for the Hénon and Heiles system when a larger Δt leads to slower convergence and also the assimilation interval at which convergence may cease altogether. The application to a simplified low order shallow water model describing coupled Rossby and gravity waves and with a forced-dissipative perturbation extends the previous result to systems which are a more realistic model for the atmosphere and the ocean. Formally, the same behaviour still holds when a realistic dissipation scheme is applied with increasing amplitudes or when strongly dissipative systems, which are not forced-dissipative perturbations of Hamiltonians, are used.

1. Introduction

One of the main objectives in geophysical data assimilation is to use all the accessible information (observations, dynamics) of the current state of the atmosphere, and perhaps in future for the ocean, in order to numerically predict the future state. Data assimilation has evolved during the last two decades or so and is now often explicitly formulated to find the initial conditions for a forecast model run. As observations are always noisy, a filtering process has to be introduced. Two equivalent optimal approaches were developed to deal with this problem; sequential estimation based on the Kalman filter (Kalman, 1960; Jazwinski, 1970), where observational data are treated recursively, and the variational or global optimization problem, based on a minimization of the distance between a certain 4D model trajectory and the observed data, used for the first time by Sasaki (1958). With large computers, both optimal techniques are being explored more and more for real data assimilation problems, although as yet forecasting centers only use them in a simplified form (Courtier et al., 1994) because the computational load is too great.

The variational approach is currently favoured at forecasting centres and it has lead to the introduction of the adjoint model technique (Lewis and Derber, 1985 and LeDimet and Talagrand, 1986). Assimilation by sequential data insertion, aiming at converging towards the true trajectory of the system, has, due to its ease of application, been used for most meteorological data assimilation during the last two decades or

* Corresponding author.

so (Charney et al, 1969; Williamson and Kasahara, 1971; Talagrand, 1981a; and many others) and has also been applied in oceanography (Cooper and Haines, 1996 and references therein). If sequential assimilation includes a proper forward prediction of error fields this represents a Kalman filter; the Kalman filter approach can be viewed as a series of model runs with successively updated initial conditions. It is important to understand how the application of sequential techniques may lead to convergence of the modelled system to the observations and this is the topic of this paper. Periodic updating is in use today in many centers, e.g., the National Center for Environmental Predictions (NCEP, formerly the National Meteorological Center) Global Data Assimilation System (GDAS) applies updating by periodically replacing the vorticity while leaving undisturbed the horizontal divergence which appears to satisfactorily converge due to model dynamics (Browning and Kreiss, 1996).

Williamson and Dickinson (1972) used periodic updating in a simplified 1D linear shallow water model. They suggested that the updating interval should be greater than the decay time of inertia-gravity waves in order to avoid inserting redundant data in a dissipative system. Talagrand (1981b) (hereafter T81b) also showed, using the traditional tangent linear system, that a better convergence could be obtained for larger updating time interval, in a forward-backward assimilation scheme using the shallow water model. If the issue were only one of data redundancy then there would be little interest in this point however we will show here that slower convergence when more data is used is a common feature of the sequential assimilation procedure when the data assimilation frequency becomes high enough.

Simple dynamical systems are often useful because they provide a framework to test assimilation procedure. The Lorenz model, for example, (Lorenz, 1963) has been extensively used for this purpose (Miller et al, 1994 and Gauthier, 1992, and others) because it represents a modal truncation of the Navier-Stokes equations in the Bénard problem, and because its chaotic attractor exhibits regime-like behaviour. Real geophysical systems often behave as conservative systems on short timescales but as forced/dissipative systems on longer timescales. Since we are interested in the convergence of the assimilation procedure in relation to the updating frequency, a simple chaotic Hamiltonian system is studied. In obtaining an explanation of its behaviour under sequential data assimilation some more generally applicable results are derived which apply to all Hamiltonian systems and to a lesser extent to dissipative systems. We will present the model in Section 2. Sequential assimilation methodology and application is presented in Section 3. Analysis of the results is in Section 4. Extension of the analysis to a simplified low order shallow water model is presented in Section 5 and summary and conclusions are in the last section.

2. A near-integrable Hamiltonian system

2.1. System introduction

Conservative or Hamiltonian systems constitute a class of dynamical systems very often encountered in physics, astronomy and meteorology. Most geophysical problems can usually be regarded as Hamiltonian on some timescale, after which forcing and dissipative processes must be considered. A Hamiltonian formulation of geophysical fluid dynamics can be found in Shepherd (1990). Hamiltonian systems also arise naturally in subjects such as calculus of variations or optimal control theory (see for example Lawden, 1975). The example focused on in this manuscript is a two degree of freedom Hamiltonian system discussed by Hénon and Heiles (1964) (referred to as HH in the text). It models the motion of a particle in an axisymmetric gravitational potential field which reduces the motion to a plane. In HH the potential;

$$U(x, y) = \tfrac{1}{2}(x^2 + y^2 + 2x^2 y - \tfrac{2}{3}y^3) \tag{1}$$

was used in order to study the existence of a third isolating integral, i.e. to study the integrability of the system with the corresponding Hamiltonian $H(x, y, p, q)$ (or total energy);

$$H = \tfrac{1}{2}(x^2 + y^2 + p^2 + q^2) + x^2 y - \tfrac{1}{3}y^3. \tag{2}$$

Fig. 1 shows the lines of equipotential and illustrates the variation of the potential well with respect to the x and y canonical positions. The potential well behaves like a near harmonic potential for x and y within a small neighbourhood of the origin. In (2), $p = \dot{x}$ and $q = \dot{y}$ are the conjugate momenta and we let $X = (x, y)^{\mathrm{T}}$ be the vector

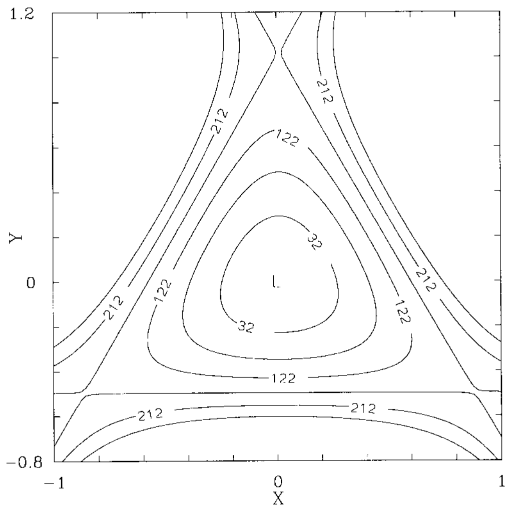

Fig. 1. The potential well $U(x, y)$ (1) of HH system. Closed equipotentials (bounded motion) are delimited by the triangle whose sides correspond to the limiting potential $U_l = 1/6$ and correspond to the separatrices. Labels scaled by 1000.

position and $V = (p, q)^T$ the vector momentum of the particle, where the superscript, T, is the transpose operator. The equations of motion are,

$$\ddot{X} = -\nabla U. \tag{3}$$

The non-integrability of this system is discussed in the literature, refer, for example, to Holmes (1982) or to Lichtenberg and Lieberman (1983). Fig. 1 shows that the particle is trapped in the well if H is less than the limiting potential energy $U_l = 1/6$, found at the singular saddle point (0, 1). The corresponding isolines form the separatrices of the system.

2.2. *Phase space representation*

The existence of H as a global invariant confines the trajectory within a three dimensional (x, y, q) volume given by;

$$U(x, y) + \tfrac{1}{2}q^2 \leqslant H. \tag{4}$$

Within this volume the trajectory may evolve periodically, quasi-periodically (regular trajectories) or unevenly (irregular behaviour). Due to the existence of this bounded domain the structure of the phase space can be studied using Poincaré maps obtained by successive intersections of the trajectory with a plane cut through the bounded volume. Poincaré maps, within the (y, q) surface with $x = 0$, have been computed for different energy levels, and some examples are shown in Fig. 2, where the different trajectories are obtained for several different initial conditions $(x(0) = 0, y(0), q(0))$ and at three different energy levels; $H = 0.0835$ (2a), $H = 0.125$ (2b) and $H = 0.165$ (2c). Fig. 2 shows that as the energy level increases the phase space progressively looses its regular (or ergodic) trajectories (2a) which are running on toroidal surfaces known as KAM (Kolmogorov, Arnold and Moser) surfaces (see for example Ott, 1994) in favour of more irregular or chaotic ones (Figs. 2b and 2c).

The irregular trajectories, generating what is known as regions of stochasticity, are known to fill a finite portion of the energy surface in phase space (Figs. 2b and 2c) and are called sometimes stochastic trajectories (see Lichtenberg and Lieberman, 1983) despite the deterministic character of the system. To see the influence of the stochastic regions on system trajectories, Fig. 3 shows the error growth or the divergence of initially nearby trajectories, for the Fig. 2 energy levels. Fig. 3 shows clearly the exponential initial error growth for chaotic trajectories. It also shows that, even within the chaotic trajectories family, the divergence is stronger for trajectories with higher energy.

3. Sequential data assimilation

Assimilation by sequential insertion is tested using a twin experiment format with the assumption of a perfect model. For each experiment, the system (3) is integrated forward in time with a specified energy level and initial conditions. The model trajectory is saved each time step and will be treated as (perfect) observations (or truth). The case of noisy observations is not treated here. The assimilation model is initialised with a random set of initial conditions and sequential data assimilation is performed by updating some of the model variables at regular intervals. At each assimilation time the observed variables are updated with the truth data, while the remaining variables are kept unchanged (thus knowledge of model variable correlations, which could be included in the update, are neglected). As the model is conservat-

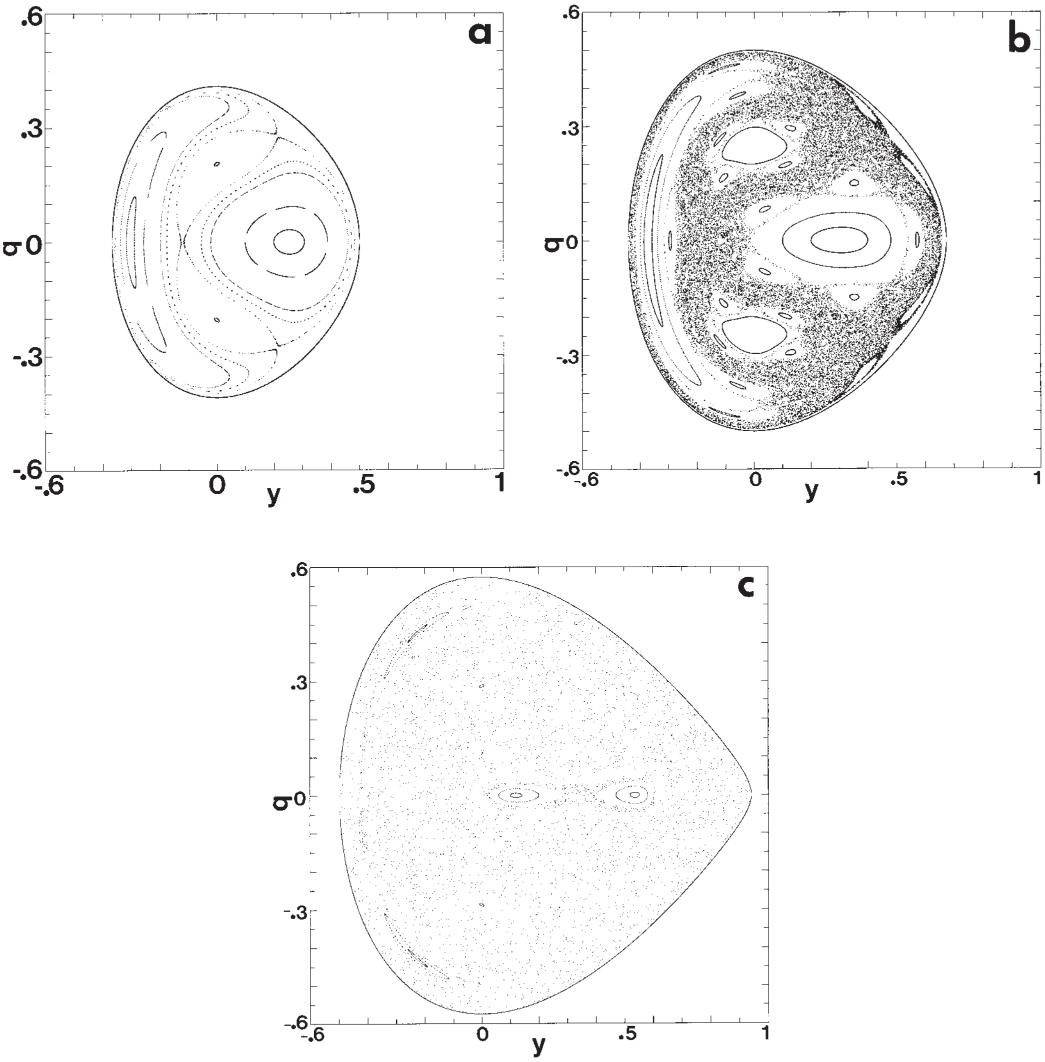

Fig. 2. Poincaré map within the (y, q) plane showing the intersection of trajectories with the plane of section at energy levels: $H = 0.0835$ with six trajectories whose initial conditions $(y(0), q(0))$ are $(0.1, 0)$, $(0.2, 0)$, $(0.02, 0.08)$, $(0, 0.2)$, $(0.2, 0.25)$ and $(-0.176, 0.)$ (a); $H = 0.125$ with one more trajectory with $(0.1, 0.2)$ as initial condition (b) and $H = 0.165$ with only five trajectories with initial conditions $(0.1, 0)$, $(0.2, 0)$, $(0, 0.29)$ $(0.28, 0)$ and $(0.1, 0.2)$ (c). In all initial conditions $x(0) = 0$. Notice the stochastic area between the KAM tori filled with one trajectory in (b). In (c) the first three produce regular trajectories, the fourth one is chaotic confined around the first two trajectories and the last one produces a chaotic trajectory filling the whole remaining volume. The time step dt is 5×10^{-3}.

ive, when new data is incorporated the energy of the system will suddenly change. For this reason, we call the procedure "non-conservative" assimilation. Other forms of "conservative" assimilation exist but are not treated here.

In the first two experiments the truth energy level is 0.125 and the truth initial conditions are quasi-periodic $(x, y, q) = (0, 0.1, 0.2)$ and chaotic $(x, y, q) = (0, -0.176, 0)$ see Fig. 2b. In order to conserve the energy numerically, a 4th-order Runge-Kutta integration scheme is used, with a time step d$t = 5 \times 10^{-3}$. We suppose that data

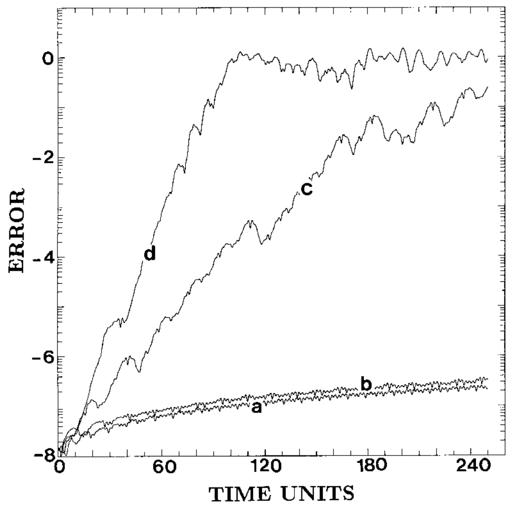

Fig. 3. Error growth as measured by the divergence of initially nearby trajectories at energy levels 0.0835 (a), 0.125 (b, c) and 0.165 (d). The initial conditions are; ((0.1, 0); $(0.1+\varepsilon, \varepsilon)$) (a, b); ((−0.176, 0); $(-0.176-\varepsilon, \varepsilon)$) (c) and ((0.1, 0.2); $(0.1+\varepsilon, 0.2+\varepsilon)$) (d) with $\varepsilon = 10^{-8}$ and time step $dt = 5 \times 10^{-3}$. The trajectories used to produce curves (a) and (b) are quasi-periodic (regular) and chaotic (irregular) for (c) and (d). Vertical axis is logarithmic.

from this run is available every Δt, which is a multiple of the time step, i.e. $\Delta t = N\ dt$, where N is an integer. Assimilation of both x and y variables is performed for different intervals, Δt, between observations. Fig. 4 shows the graph of the rms errors of the model variables for different updating intervals for a quasi-periodic trajectory (a), and a chaotic trajectory (b). For both types of trajectory the errors decrease in all assimilation cases, however the convergence is more rapid for larger time intervals, Δt, between observations, up to a value $\Delta t_{\max}$ with $N = 250$–300, after which the rate of convergence decreases until no convergence occurs.

At higher energy levels, near the separatrices of Fig. 1, the same phenomenon is observed. Fig. 5 shows the graph of the assimilation rms error for a chaotic trajectory $(x, y, q) = (0, 0.1, 0.2)$ at the energy level 0.165. The same basic result holds for quasi-periodic trajectories at this energy (not shown) and also when we assimilate p and q or x and q (Fig. 6). With non-conservative assimilation the energy is recovered at the end of the process (Fig. 7). The horizontal segments in Fig. 7b correspond to the energy conservation during integration of the system, with the jumps corresponding to updating.

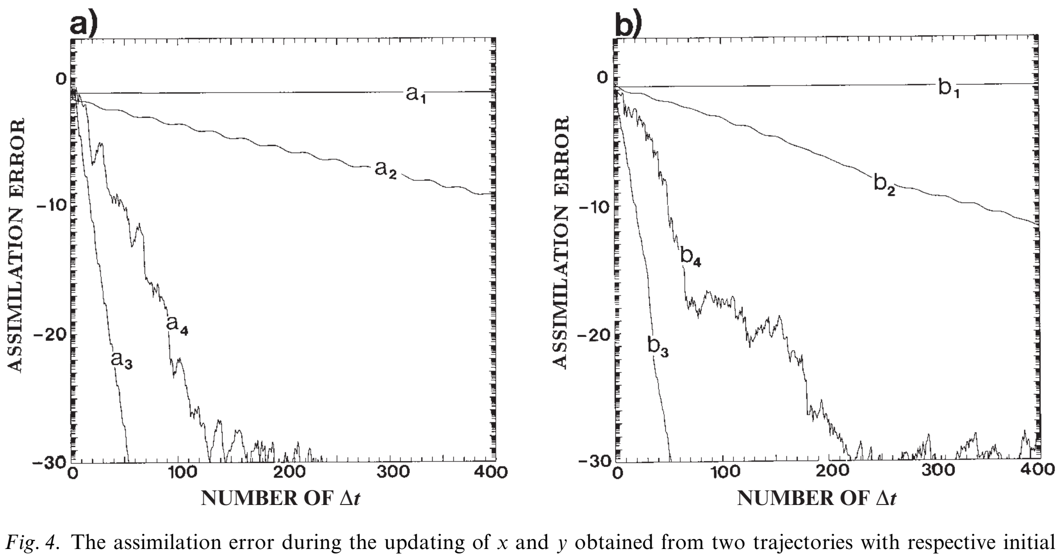

Fig. 4. The assimilation error during the updating of x and y obtained from two trajectories with respective initial conditions (0.1, 0.2) (a) and (−0.176, 0) (b) at energy $H = 0.125$ and for different updating time intervals Δt fixed to 2dt (a1, b1), 50dt (a2, b2), 250dt (a3, b3) and 1000dt (a4, b4) with $dt = 5 \times 10^{-3}$. The trajectories (a) and (b) are respectively quasi-periodic and chaotic. Vertical axis is logarithmic. Horizontal axis is number of Δt.

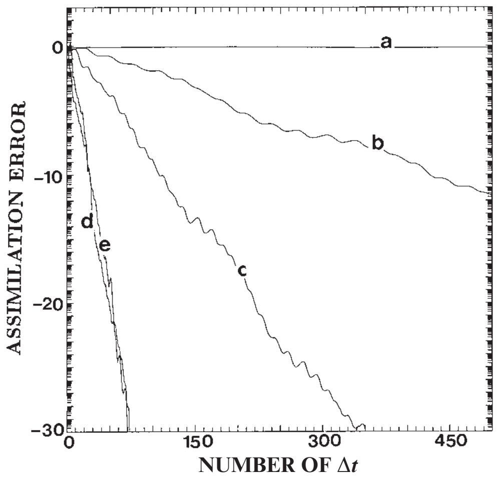

Fig. 5. As in Fig. 4a, but at $H = .165$ and Δt fixed to 5dt (a), 50dt (b), 100dt (c), 250dt (d) and 300dt (e). Notice the limit $\Delta t_{\max} \approx 250 - 300dt$ for increasing convergence with Δt. Axes as Fig. 4.

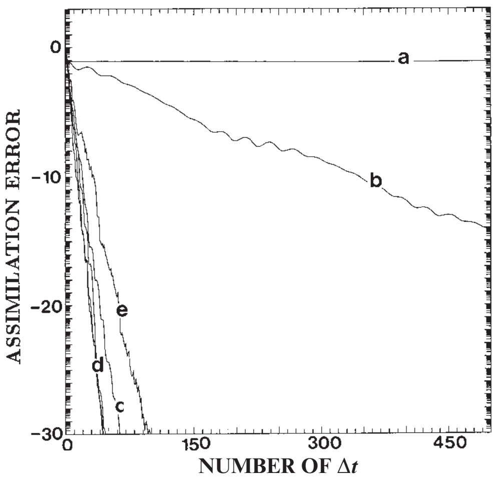

Fig. 6. As in Fig. 4b, but where x and $\dot{y}$ are updated and with Δt fixed to 2dt (a), 50dt (b), 200dt (c) and 500dt (e). The curves corresponding to 250dt, 300dt and 350dt are labelled altogether (d) showing that $\Delta t_{\max} \approx (250 - 350)$ dt (see text). Axes as Fig. 4.

The above results seem to be rather surprising because they show that more data assimilation can actually lead to a slower convergence of the dynamical system. This is not a question of data redundancy as in Williamson and Dickinson (1972). It is very important to understand why this can occur in case we ever faced the same behaviour in an important geophysical system such as in those used for atmospheric data assimilation. To understand this phenomenon we need to perform an analysis of the convergence process.

4. Analysis

4.1. Theory of assimilation convergence

Data assimilation results obtained by Williamson and Kasahara (1971) showing the convergence of the sequential updating assimilation process in a General Circulation Model (GCM) were interpreted by Williamson and Dickinson (1972) using linear perturbation theory of the updating process. Talagrand (1981a) studied the conditions of convergence of data assimilation in a linearized shallow-water model and pointed out that nonlinearity can accelerate the convergence. Talagrand (1981b) derived a convergence criterion, based on stability analysis, for a forward-backward assimilation over a particular time interval (a 4DVar pre-cursor) which depended on the spectral radius of the amplification matrix. In continuous sequential assimilation, the amplification matrix between assimilations is a function of the time, hence a general criterion of convergence is more difficult to find, but a similar approach can be used. The problem of sequential assimilation is considered below.

Consider a general m-dimensional dynamical system (where $m = 4$ for system (3));

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X), \tag{5}$$

where $X = (x_1, ..., x_m)^{\mathrm{T}}$ is an m-component vector and $F(X) = (F_1(X), ..., F_m(X))^{\mathrm{T}}$ is an m-valued function of X, where for a given $\tilde{X}$, at t_0, (5) has a unique solution or trajectory $X(\tilde{X}, t)$. The system vector at time t_i is denoted $X(t_i)$. For system (3), $X = (x, y, p, q)^{\mathrm{T}}$ and $F(X) = (p, q, -x - 2xy, -y - x^2 + y^2)^{\mathrm{T}}$.

Any small perturbation, δX_0, superimposed upon the trajectory at time t_0, will evolve according to the tangent linear equation,

$$\frac{\mathrm{d}\delta X}{\mathrm{d}t} = F'(X(t))\delta X, \tag{6}$$

where $F'(X(t)) = \partial F(X(t))/\partial X$, or simply $F'(t)$, is a second order tensor representing the Jacobian of F at $X(t)$. The solution $\delta X(t)$ of (6) at time t,

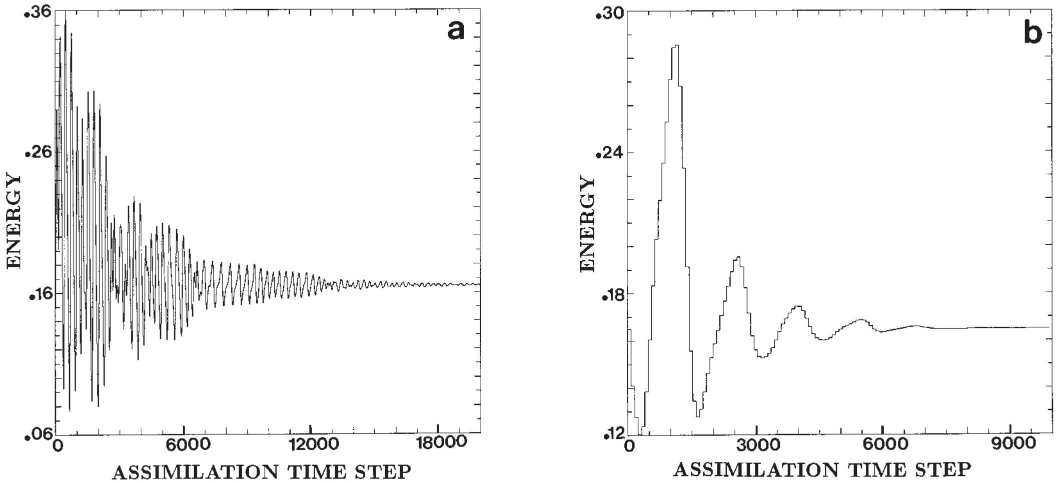

Fig. 7. Energy convergence during the non-conservative assimilation process (x, y updating) for the chaotic trajectory of Fig. 5 with Δt fixed to $5dt$ (a) and $80dt$ (b), respectively. The horizontal segments in (b) indicate the conservation the energy between the assimilation times.

given $\delta X(t_0)$ at time t_0, is given by,

$$\delta X(t) = R(t, t_0)\delta X(t_0), \tag{7}$$

where the operator R is the resolvent (also called the propagator) of system (6). The resolvent satisfies the following two important relationships;

$$\frac{\partial}{\partial t} R(t, t_0) = F'(t)R(t, t_0), \qquad \forall\, t, t_0, \tag{8}$$

and obviously;

$$R(t, t) = I_m, \qquad \forall\, t, \tag{9}$$

where I_m is the $m \times m$ identity operator. Eq. (8) can be obtained by differentiation of eq. (7) with respect to time. The resolvent is not, generally, easily accessible, however for any time t_0 and infinitesimal Δt, one can use the Taylor expansion:

$$R(t_0 + \Delta t, t_0) = I_m + \Delta t \frac{\partial}{\partial t} R(t, t_0)|_{t=t_0} + \dots + \frac{\Delta t^n}{n!} \frac{\partial^n}{\partial t^n} R(t, t_0)|_{t=t_0} + \mathrm{O}(\Delta t^{n+1}). \tag{10}$$

At $\mathrm{O}(\Delta t^3)$ the previous expansion, together with relations (8) and (9) give,

$$R(t_0 + \Delta t, t_0) = I_m + \Delta t F'(t_0) + \frac{\Delta t^2}{2} (\dot{F}'(t_0) + F'^2(t_0)) + \mathrm{O}(\Delta t^3), \tag{11}$$

where

$$\dot{F}'(t) = \frac{\mathrm{d}}{\mathrm{d}t} F'(t) = \frac{\partial^2 F}{\partial X^2} F(X(t)) = F''F. \tag{12}$$

In (12), $F'' = \partial^2 F/\partial X^2$ is a third order tensor with components, $(\partial^2 F/\partial X^2)_{ijk} = \partial^2 \mathrm{F}_i/\partial \mathrm{x}_i \partial \mathrm{x}_k$, and $\dot{F}'(t)$ is the operator (matrix) obtained by contraction of F'' with F, whose components are,

$$(\dot{F}'(t))_{ij} = \sum_{k=1}^{m} \frac{\partial((F')_{ij})}{\partial x_k} F_k = \sum_{k=1}^{m} \frac{\partial^2 F_i}{\partial x_k \partial x_j} F_k. \tag{13}$$

For assimilation purposes, system (5) is split into two parts;

$$\begin{cases} \dot{X}_\mathrm{a} = F_\mathrm{a}(X) \\ \dot{X}_\mathrm{u} = F_\mathrm{u}(X) \end{cases} \tag{14}$$

with $X = (X_\mathrm{a}, X_\mathrm{u})^\mathrm{T}$, where X_a and X_u stand respectively for the assimilated and unchanged variables in the assimilation procedure. Accordingly, the resolvent of (7) is then written as:

$$R(t, t_0) = \begin{pmatrix} R_{\mathrm{aa}}(t, t_0) & R_{\mathrm{au}}(t, t_0) \\ R_{\mathrm{ua}}(t, t_0) & R_{\mathrm{uu}}(t, t_0) \end{pmatrix}. \tag{15}$$

At the assimilation times t_k, the error, $\delta X_\mathrm{a}(t_k)$, of the assimilated variables is set to zero as observations are made, while the error for the remaining variables, $\delta X_\mathrm{u}(t_k)$, is kept unchanged so that the error model (7) is then re-initialised with errors $(0, \delta X_\mathrm{u}(t_k))^\mathrm{T}$. The error at the next assimilation time t_{k+1}, $\delta X(t_{k+1})=(\delta X_{\mathrm{a}}(t_{k+1}),\ \delta X_{\mathrm{u}}(t_{k+1}))^{\mathrm{T}}$, is given by (7). But again, $\delta X_{\mathrm{a}}(t_{k+1})$ is set to zero at the assimilation time and $\delta X_{\mathrm{u}}(t_{k+1})$ is kept unchanged. The convergence of the assimilation process is then linked to the amplification matrix, $R_{\mathrm{uu}}(t_k, t_{k-1})$, in the sequence,

$$\delta X_{\mathrm{u}}(t_k) = R_{\mathrm{uu}}(t_k, t_{k-1})\delta X_{\mathrm{u}}(t_{k-1}). \tag{16}$$

In T81b, the amplification matrix for a forward-backward assimilation over a specified small time interval is calculated once only and the convergence process turns out to be an explicit function of the spectral radius of this matrix. In the continuous sequential assimilation scheme (16), the amplification matrix is a function of time and we have to deal with it according to the situation arising. If the matrix $R_{\mathrm{uu}}(t_k, t_{k-1})$ has a simple form, e.g., triangular, with a spectral radius, r_k, then the corresponding radius product, $\prod_{k=1}^{n} r_k$, will determine the convergence. Normally we must consider the sequence matrix product,

$$A_n = \prod_{k=1}^{n} R_{\mathrm{uu}}(t_k, t_{k-1}), \tag{17}$$

whose spectral radius will be denoted $\rho_n = \rho(A_n)$.

Although the spectral radius function itself is not a matrix norm (nilpotent matrices are counter-examples), for any matrix M, it is the greatest lower bound for all the matrix norms $\|M\|$, of M (see for example Horn and Johnson, 1988). This means that;

$$\rho(M) = \operatorname{Inf}\{\|M\|, \|\cdot\| \text{ is a matrix norm}\}.$$

Therefore a necessary condition for (17) to converge to the null matrix, O, is that,

$$\rho_n = \rho(A_n) \to 0$$

as n goes to infinity.

4.2. *Application to Hamiltonian systems*

For the HH system we have:

$$F'(t) = \begin{pmatrix} 0 & I_2 \\ \begin{pmatrix} -1-2y & -2x \\ -2x & -1+2y \end{pmatrix} & 0 \end{pmatrix}, \tag{18}$$

and

$$F''F(t) + F'^2(t) =$$
$$\begin{pmatrix} \begin{pmatrix} -1-2\mathrm{y} & -2\mathrm{x} \\ -2\mathrm{x} & -1+2\mathrm{y} \end{pmatrix} & 0 \\ \begin{pmatrix} -2\mathrm{q} & -2\mathrm{p} \\ -2\mathrm{p} & 2\mathrm{q} \end{pmatrix} & \begin{pmatrix} -1-2\mathrm{y} & -2\mathrm{x} \\ -2\mathrm{x} & -1+2\mathrm{y} \end{pmatrix} \end{pmatrix}. \tag{19}$$

It is clear from (11), (15) and (18) that the R_{uu} component of the resolvent has no $\mathrm{O}(\Delta t)$ terms when X_{u} equals (x, y) or (p, q). What is more it is clear that for any Hamiltonian system satisfying (3), the convergence of the assimilation process, when assimilating all the canonical (position) variables (x and y here) or all the conjugate (momentum) variables (p and q here), is of order $\mathrm{O}(\Delta t^2)$, i.e. the amplification matrix is of order $\mathrm{O}(\Delta t^2)$ at least for small time intervals Δt. Furthermore, from (19), the assimilation of the canonical (position) or conjugate (momentum) variables give the same convergence rate at $\mathrm{O}(\Delta t^2)$, due to the symplectic form of Hamiltonian systems.

Mixing the variables by assimilating either (y, p) or (x, q) also produces convergence at $\mathrm{O}(\Delta t^2)$ for the HH system. From (11), (18) and (19) the $O(\Delta t^2)$ amplification matrix, when (y, p) are updated, is given by;

$$R_{\mathrm{uu}}(t_n, t_{n-1}) =$$
$$\begin{pmatrix} 1-\dfrac{\Delta t^2}{2}(1+2y_n) & 0 \\ -2x_n\Delta t - p_n\Delta t^2 & 1-\dfrac{\Delta t^2}{2}(1-2y_n) \end{pmatrix},$$

where $(x_n, y_n, p_n, q_n) = (x(t_n), y(t_n), p(t_n), q(t_n))$. For this case the triangular form means that the spectral radius is explicitly known and,

$$\rho_n = \max(\rho_n^+, \rho_n^-) \tag{21}$$

where,

$$\rho_n^+ = \prod_{k=1}^{n} \left|1-\frac{\Delta t^2}{2}(1+2y_k)\right|,$$

$$\rho_n^- = \prod_{k=1}^{n} \left|1-\frac{\Delta t^2}{2}(1-2y_k)\right|,$$

so that ρ_n tends to zero, for small Δt when n goes to infinity. The limit of ρ_n^+ is straight-forward, since y_k belongs to the interval $[-1/2, 1]$, but it is slightly more complicated for ρ_n^- because $(1-2y_k)$ is sometimes negative, and we require the asymptotic behaviour for large n. Appendix A shows that for large n, y_k can be replaced with $\bar{y}$, the mean trajectory value which is close to zero for all bound trajectories, i.e., at the bottom of the potential well in Fig. 1. Appendix A also shows that $\rho_n(\Delta t)$ is a decreasing function of Δt in the range $\Delta t^2 < 2$, i.e.,

$$\rho_n(\Delta_2 t) \leqslant \rho_n(\Delta_1 t) \qquad \text{for } \Delta_1 t \leqslant \Delta_2 t \leqslant \sqrt{2},$$

where $\Delta_1 t$ and $\Delta_2 t$ correspond to two different time intervals between assimilating observations from the original trajectory. For $\Delta t^2 > 2$, the decreasing process is expected to reverse. This limiting value of Δt leading to better convergence is in approximate agreement with the value obtained from Fig. 6, where for $\Delta t \leqslant (250\text{–}300)\,\mathrm{d}t$ (with $\mathrm{d}t = 5 \times 10^{-3}$) the assimilation convergence is improving with Δt. Fig. 8 indicates the variation of ρ_n given by (21) as a function of Δt for a chaotic trajectory at energy level 0.165. It shows clearly the increasing convergence for increasing time periods Δt.

The more natural way to compare the convergence rates is to compare two different assimilation cycles, n_1 and n_2, $(n_2 < n_1)$ for which:

$$n_1 \Delta_1 t = n_2 \Delta_2 t = \sigma.$$

In this case, with large n_1 and n_2 and relatively small $\Delta_1 t$ and $\Delta_2 t$ we still have $\rho_{n_2}(\Delta_2 t) \leqslant \rho_{n_1}(\Delta_1 t)$, which is an important result because the $n_2 \Delta_2 t$ cycle actually assimilates less data. This is an immediate consequence of the asymptotic behaviour of ρ_n given in (A5) (Appendix A) which leads to,

$$(\log \rho_{n_2} - \log \rho_{n_1}) \sim (\Delta_1 t - \Delta_2 t)\sigma(1 - 2|\bar{y}|) \leqslant 0. \tag{22}$$

This is well illustrated in Fig. 9 which shows the spectral radius ρ_n of A_n, when (x, q) are updated, as a function of the updating time interval Δt and the number of assimilation cycles, n. Hyperbolae with $n\Delta t = \sigma$ are also shown. The calculations have been performed to an accuracy $O(\Delta t^3)$ where the third order term in expansion (10) was included. Notice the minimum of ρ_n occurs at $\Delta t \approx 1.36$ time units which is close to the value of $\sqrt{2}$ predicted by the theory.

We now return to the more complicated case. When assimilating (x, y) or (p, q), the $O(\Delta t^2)$ amp-

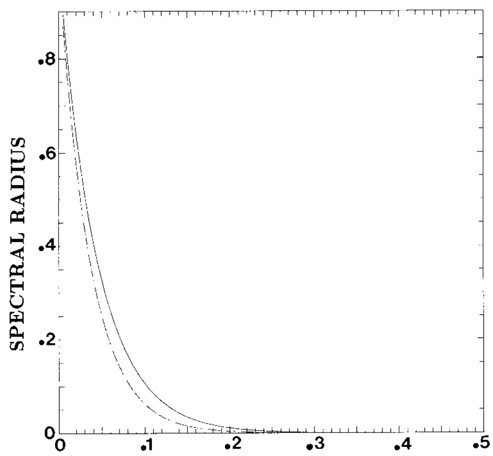

Fig. 8. Variation of the spectral radius ρ_n of the amplification matrices product given by (21) as a function of the assimilation time interval Δt and computed from the chaotic trajectory of Fig. 5. The dashed line is for ρ_n^- and the solid line for ρ_n^+. The first 10^4 time steps were considered.

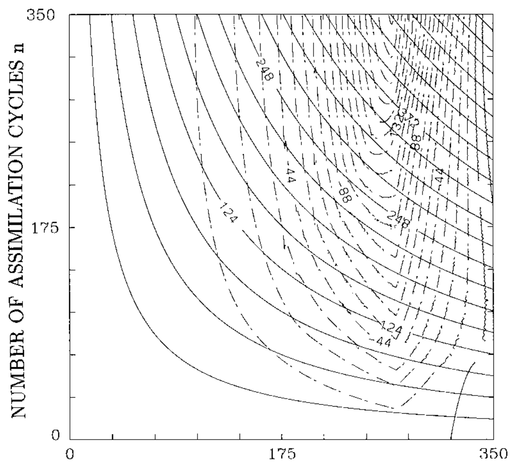

Fig. 9. The logarithm of the spectral radius, $\log \rho_n$, of A_n as a function of the updating time interval Δt and the number of assimilation cycles n performed when updating (x, q) for the experiment $H = 0.165$ and $(y(0), q(0)) = (0.1, 0.2)$ using the expansion (11) of the resolvent at $O(\Delta t^3)$ (dashed lines) and the hyperbolas $n\Delta t = \sigma$ (solid lines).

lification matrix becomes,

$$R_{\mathrm{uu}}^{n} = \begin{pmatrix} 1 - \dfrac{\Delta t^2}{2}(1 + 2y_n) & -\Delta t^2 x_n \\ -\Delta t^2 x_n & 1 - \dfrac{\Delta t^2}{2}(1 - 2y_n) \end{pmatrix}. \tag{23}$$

For the HH run, the spectral radius ρ_n of A_n, from (17) and (19), is calculated and plotted as a function of Δt, in Fig. 10. It also shows the increase of convergence with time interval Δt, as a decrease in ρ_n, up to an interval $\Delta t \sim (250 - 300)\,\mathrm{d}t$, i.e., around 1.375 time units, which agrees well with the assimilation errors reported in Figs. 4, 5 and with the predicted value $\Delta t = \sqrt{2}$.

The theory in Appendix A also predicts a limit on Δt after which convergence of the sequential assimilation may cease. Generally convergence is predicted for $\Delta t^2 \leqslant 4$. Table 1 gives the maximum values of Δt, for different trajectories at different energy levels, where convergence of the assimilation (in the full system) no longer occurs. These values are all greater than the theoretically predicted value of $\Delta t = 2$ (Appendix A). However nonlinearity can generally improve the convergence and this allows convergence for periods Δt larger than those predicted from the $\mathrm{O}(\Delta t^2)$ expansion of the resolvent. Similar conclusions were derived empirically by Williamson and Kasahara (1971), and have also been noted in T81b.

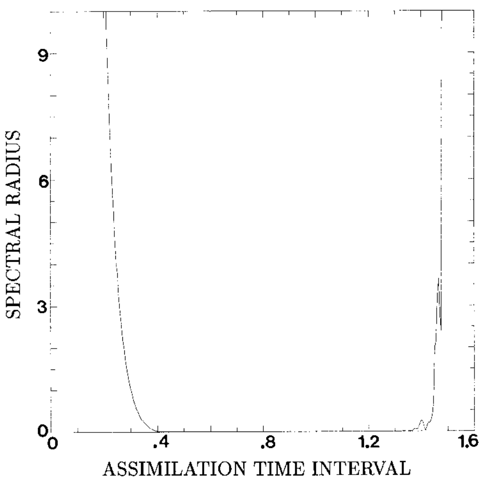

Fig. 10. As in Fig. 8, but the spectral radius ρ_n of the product A_n is entirely calculated from (17) and (23). Notice the assimilation time interval Δt around 1.38 where the convergence process is no longer increasing with Δt. Vertical axis scaled by 10^3.

Table 1. *Approximate maximum values of the assimilation time interval Δt (time units) for which convergence is no longer obtained*

Initial conditions (x, y, q)	Energy levels H		
	$H = 0.0835$	$H = 0.125$	$H = 0.165$
(0, 0.1, 0.2)	5.80	2.90	2.35
(0, −0.176, 0)	6.25	3.80	2.40
(0, 0.1, 0)	6.45	3.95	2.85
(0, 0, 0.28)	6	2.95	2.125
(0, −0.2, −0.25)	6.50	5.85	2.525

As real geophysical systems such as the atmosphere and oceans are almost Hamiltonian on short timescales, the natural question is whether these results also hold for weakly forced/dissipative systems which are Hamiltonian to first order. This has been addressed in the next section.

5. Application to simplified geophysical systems

5.1. Forced-dissipative perturbation of Hamiltonian systems

In order to address these issues, a simplified 5-component primitive equations (PE) model derived by Lorenz (1986) (hereafter L86) has been used. L86 derived the following simplified PE of a (three Fourier mode) spectral shallow water model. The system reads:

$$\begin{cases} \dot{x}_1 = -x_2 x_3 \\ \dot{x}_2 = x_1 x_3 \\ \dot{x}_3 = -x_1 x_2 - b x_4 \\ \dot{x}_4 = b x_3 - x_5 \\ \dot{x}_5 = x_4. \end{cases} \tag{24}$$

System (24) was derived from a 9-component spectral shallow water model (Lorenz, 1980) and describes the interaction between the rotational component of the wind, associated with three Fourier coefficients x_1, x_2, x_3, and a single gravity wave mode associated with the divergence coefficient x_4, and the geopotential associated with the coefficient x_5. In (24) b is the ratio of the wave length of mode 3 to the Rossby circumference of deformation and is fixed to $b = 0.5$. We write system (24) in vector form as:

$$\frac{dX}{dt} = G(X), \tag{25}$$

where $X = (x_1, x_2, x_3, x_4, x_5)^T$ and $G(X) = (-x_2x_3, x_1x_3, -x_1x_2 - bx_4, bx_3 - x_5, x_4)^T$. L86 transformed (25) into a normal-mode form which was used later by Bokhove and Shepherd (1996) who transformed it into a canonical Hamiltonian form to study the slowest invariant manifold, but in this manuscript we use the original system (25). This system exhibits multiple timescales (Rossby and gravity timescales). The influence of multiple timescales on the updating convergence has been studied within a linear framework in Browning and Kreiss (1996) and is not addressed here.

System (25) can be written in the "non-reduced" Hamiltonian form:

$$\frac{dF(X)}{dt} = [F, H], \tag{26}$$

with Hamiltonian $H = \frac{1}{2}(x_1^2 + 2x_2^2 + x_3^2 + x_4^2 + x_5^2)$ and F is any function of the variable X. The skew symmetric generalized Poisson bracket [.,.] is given by

$$\begin{aligned}[A, B] = {} & \left(-x_2\frac{\partial A}{\partial x_1} + x_1\frac{\partial A}{\partial x_2}\right)\frac{\partial B}{\partial x_3} \\ & + \frac{\partial A}{\partial x_3}\left(x_2\frac{\partial B}{\partial x_1} - x_1\frac{\partial B}{\partial x_2} - b\frac{\partial B}{\partial x_4}\right) \\ & + \frac{\partial A}{\partial x_4}\left(b\frac{\partial B}{\partial x_3} - \frac{\partial B}{\partial x_5}\right) + \frac{\partial A}{\partial x_5}\frac{\partial B}{\partial x_4},\end{aligned} \tag{27}$$

and is shown to satisfy the Jacobi identity equation: $[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$. One of the important dynamical properties of the chaotic system (25) is that it has some of the common features observed in the atmosphere in the sense that it can exhibit regime-like behaviour rather similar to the Lorenz (1963) system. This is illustrated in Fig. 11 which shows the time variation of the component x_1, representing the low frequency variation of the nondivergent part of the flow in the simplified shallow water model. One notices a gravity mode with high frequency oscillations superimposed onto a low frequency Rossby mode.

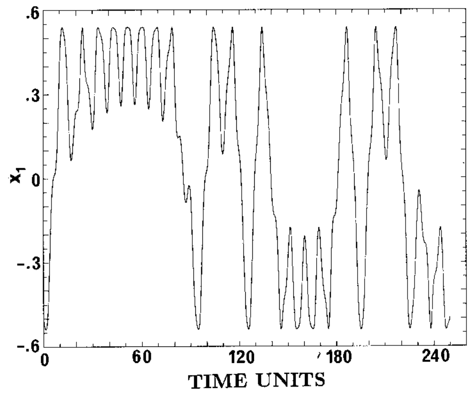

Fig. 11. Temporal evolution of the variable x_1 for a chaotic trajectory of system (25) with initial condition $(-0.5, 0.2, 0.3, 0.4, 0.4)$ and time step $dt = 5 \times 10^{-3}$.

Several twin experiments with assimilation updating are carried out for this Hamiltonian case. Fig. 12 shows the convergence of the assimilation when (x_1, x_3, x_5) are updated for different updating time intervals Δt. Again, the convergence is faster for larger Δt up to a value $\Delta t_{max} \sim (280 - 290)\,dt$ (not shown in the figure)

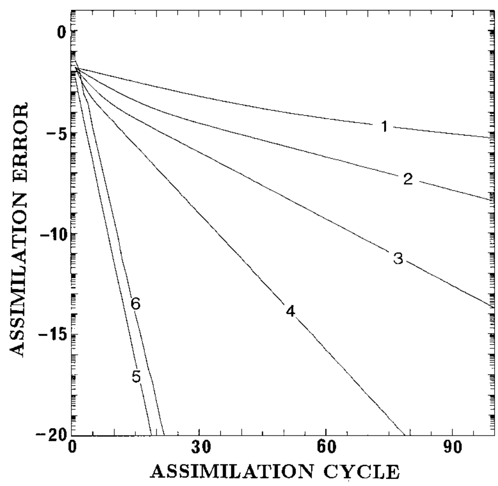

Fig. 12. Convergence of the assimilation in the L86 model when (x_1, x_3, x_5) are updated each $\Delta t = 10dt$ (curve 1), $25dt$ (curve 2), $50dt$ (curve 3), $100dt$ (curve 4), $200dt$ (curve 5) and $300dt$ (curve 6). The control run is performed with $dt = 0.005$ for 10^5 time steps with initial condition $(1, 0, -0.5, 0.5, 0.4)$ and $b = 0.5$. Vertical scale is logarithmic. Horizontal axis is number of Δt.

after which the convergence slows down for further increase of the assimilation interval until the process eventually ceases to converge. This behaviour is also observed when only two of the five variables, (x_1, x_3) or (x_1, x_5) for example, or only one variable (but not for every initial condition), are updated. The value $\Delta t_{\max}$ is again close to the value predicted from the $O(\Delta t^2)$ expansion of the resolvent as shown in Appendix B. To make the system forced/dissipative, system (25) is transformed to yield:

$$\frac{dX}{dt} = G(X) - \Lambda X + F, \tag{28}$$

where the dissipation operator Λ is a 5×5 diagonal matrix with elements v_i $(i = 1, \ldots, 5)$ and where v_i is the rate of dissipation acting upon the tendency of x_i and $F = (f_1, f_2, f_3, f_4, f_5)^{\mathrm{T}}$ a constant forcing. Unfortunately the nature of the system means that any dissipation on any of the variables x_i will make the system shrink to a steady state or to a periodic limit cycle. But it is possible to weakly force/damp the system so that the collapse occurs slowly and long before this, while the system still shows "chaotic behaviour", the assimilation may already have converged.

System (25) has two quadratic invariants, $E_1 = x_1^2 + x_2^2$ and $E_2 = x_2^2 + x_3^2 + x_4^2 + x_5^2$. If only v_1 and f_1 are taken to be nonzero so that x_1 is the only forced-dissipated variable, E_1 decreases while E_2 remains invariant. Asymptotically, the system shrinks and projects onto a sphere while becoming uncoupled. There are three generic cases to consider.

(1) If all dissipated variables (in this case x_1) are also assimilated, Appendix B shows trivially that the amplification matrix remains $O(\Delta t^2)$ and the Hamiltonian like behaviour remains. To illustrate, we assimilate (x_1, x_3, x_4) with $v_1 = 0.1$ and $f_1 = 0.01$. Fig. 13a shows the assimilation convergence for $\Delta t = N\,dt$ $(N = 10, 25, 50, 100, 200)$.

(2) If dissipated variables are unassimilated but some variables are undissipated and unassimilated then the first term in the amplification matrix is now $O(\Delta t)$, and varies as $(1 - v\Delta t)$. But Appendix B shows that the convergence after n assimilation cycles, ρ_n, still varies at $O(\Delta t^2)$ due to the undissipated and unobserved variables and therefore Hamiltonian like behaviour still remains. This result is illustrated in Fig. 13b where (x_2, x_3, x_4) are updated but convergence still shows increasing convergence with increasing $\Delta t < \Delta t_{\max}$.

(3) If all unobserved variables are dissipated then both the amplification matrix and the amount of convergence formally vary at $O(\Delta t)$. Appendix B shows that in this case the convergence may be independent of (small enough) Δt. This case is considered in more detail below.

In Appendix B the case in which both x_1 and x_5 are dissipated while (x_2, x_3, x_4) are observed, is considered. It is shown that, to accuracy $O(\Delta t^3)$, the convergence ρ_n behaves as:

$$\rho_n \sim 1 - \sigma v\left(1 - \frac{\sigma v}{2}\right) - \sigma \Delta t\left(\frac{\alpha}{2} - v^2\right), \tag{29}$$

where σ plays the role of system time; $\sigma = n\Delta t$ and v is the dissipation rate and α is a function of the climatology of the trajectory. If the second term in (29) is much larger than the third term then the convergence is independent of update interval. This is because the convergence of the unobserved variables depends on system dissipation rather than the need to update the other variables more often. This is the data redundancy identified in Williamson and Dickinson (1972). The need for dissipation to ensure this result was described by Browning and Kreiss (1996). Formally, even in this case, the convergence is faster for larger Δt provided $v < \sqrt{\alpha/2}$.

The main result is that the improved convergence with less frequent updating can be quite a robust result for weakly forced/dissipative systems as well as for Hamiltonian systems. Under these conditions in which the convergence is a function of the $O(\Delta t^2)$ expansion of the resolvent, the tangent linear system cannot explain the nonlinear behaviour of the convergence at small Δt. In Subsection 5.2 we explore the behaviour of some strongly dissipative systems.

5.2. *Case of strongly dissipative systems*

In order to include strong dissipation in chaotic systems, two examples are used as prototypes of simplified dissipative systems, that of Lorenz (1963) and Rössler (1976). Expressions of the resolvent for both systems at $O(\Delta t)$ are presented in Appendix C. For the Lorenz system, it can be seen for example that when the variable y is updated, the first nonzero term in the amplification matrix, $(1 - b\Delta t)$, is $O(\Delta t)$ (Appendix C). Figs.

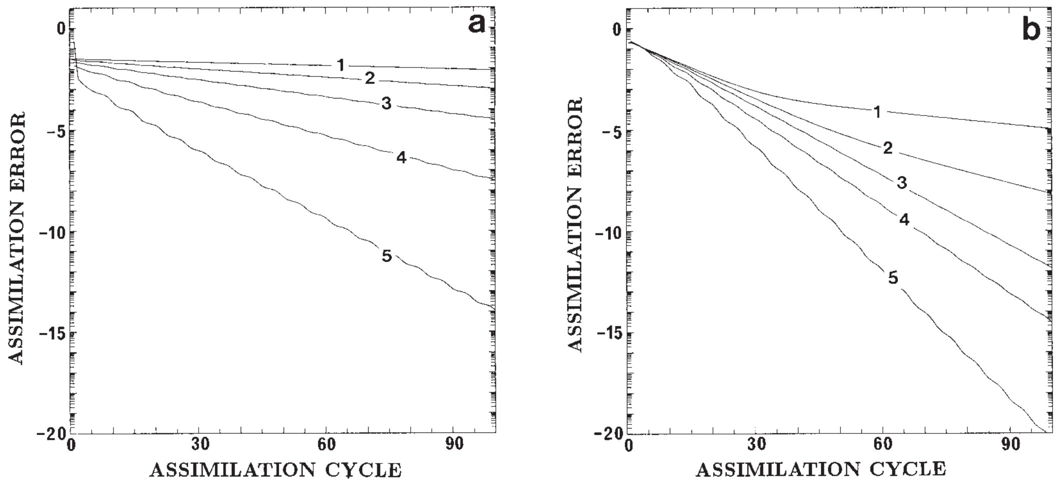

Fig. 13. As in Fig. 12, for the first 5 curves, and for the forced/dissipative perturbed system (28) when only $v_1 = 0.1$ and $f_1 = 0.01$ are nonzero and when (x_1, x_3, x_4) (a) and (x_2, x_3, x_4) (b) are updated. The control run is the same as in Fig. 11.

14a,b show the convergence of assimilation for different updating frequencies $\Delta t = N\,\mathrm{d}t$ ($N = 1, 5, 10, 20, 40$). In Fig. 14a, the parameters are the same as in the original Lorenz (1963) paper while in Fig. 14b, the parameter b was increased to $b = 10$. Although the curves in Fig. 14a are close to each other, they still show faster convergence for larger $\Delta t = N\,\mathrm{d}t$ up to the values $N = 10 - 20$ which give an approximation of the turnover point. When b is increased the curves in Fig. 14b, for $N = 1, 5, 10, 20$ become closer to each other than in Fig. 14a but, in agreement with the previous conclusions (eq. (29)) the formal increase of convergence with Δt for the first four curves is still slightly apparent.

For the Rössler (1976) system (Appendix C) we increased the dissipation parameter $\mu = 5$ and we updated the variable y each $\Delta t = N\,\mathrm{d}t$ ($N = 1, 25, 50, 100, 200$). The results, reported in Fig. 14c, clearly show, as in the Hamiltonian case, faster convergence for larger updating time interval Δt. This case highlights the importance of the $O(\Delta t^2)$ term (see the third term of (29)) in the convergence behaviour of the updating process. Finally, we also note that the increase of dissipation in the Lorenz system can increase the rate of convergence as noted by Browning and Kreiss (1996) for linear systems, but this is not a general property of dissipative systems. For example in the Rössler system (Appendix C), the updating of the variable x is not convergent, even for small Δt (since $\rho_n \sim (1 + \alpha\Delta t)^n$ at $O(\Delta t^2)$).

6. Summary and conclusions

The convergence of sequential assimilation by updating, within the framework of twin experiments in chaotic Hamiltonian systems, is the focus of this paper. The main issue is the convergence rate of the assimilation process with respect to the time interval between observations or the updating frequency. The initial study was performed using the Hénon-Heiles near-integrable Hamiltonian system which shows different phase space behaviour for different energy levels but the results and analysis method are also shown to be generally applicable to all potential-derived Hamiltonian systems.

Beginning with very frequent updating, the convergence of the process was shown to become more rapid with increasing updating time intervals Δt up to a certain point where the convergence process reverses and thereafter convergence slows as Δt increases. A $O(\Delta t^2)$ Taylor expansion of the resolvent operator of the HH system was used to explain this behaviour for small Δt, using the spectral radius of the amplification matrix products and a relationship linking the spectral radius to the matrix norms. The theory successfully pre-

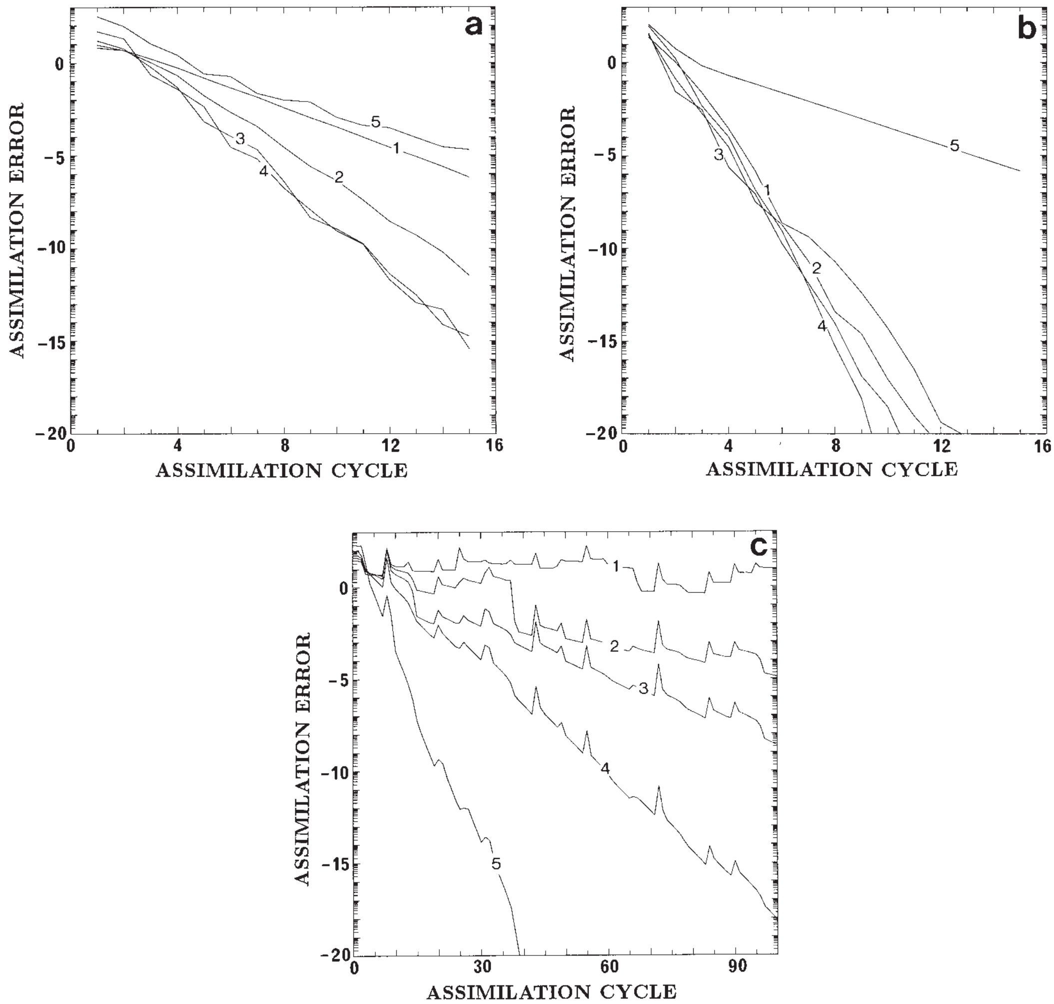

Fig. 14. The assimilation error for the Lorenz system with the original parameters (a) and with the parameter b increased to $b = 10$ (b) when the variable y is updated with $\Delta t = N\,\mathrm{d}t$, with $N = 1$ (curve 1), $N = 5$ (curve 2), $N = 10$ (curve 3), $N = 20$ (curve 4) and $N = 40$ (curve 5) and for the Rössler system with $\mu = 5$ (c) when y is updated with $N = 1$ (curve 1), $N = 25$ (curve 2), $N = 50$ (curve 3), $N = 100$ (curve 4) and $N = 200$ (curve 5). Time step is $\mathrm{d}t = 5 \times 10^{-3}$ and the vertical scale is logarithmic.

dicts the assimilation time interval for optimal convergence and also the interval at which convergence may cease. It is also shown that, due to the symplectic form of potential-derived Hamiltonian systems, the assimilation of all canonical (position) or conjugate (momentum) variables always produces this convergence behaviour for small Δt. This is because for any Hamiltonian system satisfying the equations of motion (3) (deriving from a potential), the amplification matrix, when assimilating the canonical variables (positions) or the conjugate variables (momenta), is of order $\mathrm{O}(\Delta t^2)$ and not $\mathrm{O}(\Delta t)$ for small time interval Δt.

To test the generality of this behaviour, we extended the analysis to the simplified, low-order, non-canonical, Hamiltonian primitive equations describing coupled Rossby and gravity waves in a shallow water model which was perturbed with dissipation and a constant forcing. The convergence behaviour again appears and is shown to be quite robust to small forced-dissipative perturbations. This behaviour may also occur in these systems even when the dissipation or forcing is strong. Similar results can be obtained when "natural" dissipative systems, such as those of Lorenz and Rössler, are used. This has led us to conclude that in Hamiltonian systems and their forced/dissipative perturbations, the convergence of assimilation by direct insertion is faster for larger updating time interval Δt up to the turnover limit point. For strongly dissipative systems, it is still formally possible to observe the same behaviour although data redundancy in which the convergence rate is independent of Δt, may also occur. The origin of this behaviour lies in the $O(\Delta t^2)$ dependence of the convergence rate of Hamiltonian systems during sequential data assimilation. A first order tangent linear system analysis is unable to explain this non-linear behaviour in the convergence rate. It is important to determine whether more complex systems, such as those used in Meteorological forecasting, can display this phenomenon and if there are optimal assimilation timescales associated with these systems.

7. Acknowledgments

We would like to thank the anonymous reviewers for their helpful comments and one of them for suggesting application of the Lorenz (1986) model which has considerably improved the manuscript. This work was funded under EC Environment project AGORA, CT95–0113.

8. Appendices

(A). *On the limiting behaviour of the spectral radius of the amplification matrix for system HH*

We consider the amplification matrix given by (20) obtained when (y, p) are updated. In order to analyse the spectral radius $\rho_n = \rho(A_n)$ of (21) we analyse ρ_n^+ and ρ_n^- separately. Since the variable y belongs to $[-1/2, 1]$ (actually the bounds are excluded since they correspond to the limit case when $H = 1/6$) then (Δt^2 not too large),

$$r_k^+ = \left| 1 - \frac{\Delta t^2}{2}(1 + 2y_k) \right| < 1. \tag{A1}$$

It follows that the sequence $\rho_n^+ = \prod_{k=1}^n r_k^+ = \prod_{k=1}^n |1 - \frac{1}{2}\Delta t^2(1 + 2y_k)|$goes to zero as n goes to infinity. Formally, this is because the general term, $\log r_k^+$, of the series $\sum \log r_k^+$ does not go to zero as n goes to infinity (because r_k^+ itself does not have a limit), and the series $\sum \log r_k^+$ is then divergent (to $-\infty$) and $\rho_n^+ \to 0$.

The situation is slightly different for $\rho_n^- = \prod_{k=1}^n r_k^- = \prod_{k=1}^n |1 - \frac{1}{2}\Delta t^2(1 - 2y_k)|$ since $(1 - 2y_k)$ is sometimes negative. However, for small Δt^2 we have,

$$\log \rho_n^- = \sum_{k=1}^n -\frac{\Delta t^2}{2}(1 - 2y_k) + O(\Delta t^4)$$
$$= -n\frac{\Delta t^2}{2}(1 - 2\overline{y_n}) + O(\Delta t^4), \tag{A2}$$

where $\overline{y_n} = (1/n)\sum_{k=1}^n y_k$is an average of the variable y from the trajectory. For n large enough, $\overline{y_n}$ is approximately equal to $\bar{y}$ (independent of Δt because of the regular or stochastic character of the trajectories) so that at $O(\Delta t^4)$ we have,

$$\log \rho_n^- \sim -n\frac{\Delta t^2}{2}(1 - 2\bar{y}). \tag{A3}$$

In the limiting case where $H = 1/6$, the phase space in the (x, y) plane is delimited by the triangle shown in Fig. 1. Notice that we have concentrated on the (x, y) plane because we have $(\bar{p}, \bar{q}) = (0, 0)$ (from eq. (5)). The chaotic trajectory fills the phase space and the average position can be diagnosed using the rule of triangle centres and it is close to the origin of Fig. 1. This is the case because of the geometric nature of the potential with one well (Fig. 1) and with one unique minimum located at $(x, y) = (0, 0)$. In any case we have,

$$H_1 - 2\bar{y} \geqslant 0, \tag{A4}$$

as required for convergence, which implies that $\rho_n^- \to 0$ when $n \to \infty$ (this argument can also be used for ρ_n^+).

Notice that for large n, ρ_n^+ and ρ_n^- are decreasing functions of (small) Δt. This results immediately from the asymptotic behaviour of $\rho_n^\pm$. It is possible to get an upper bound on Δt^2 for ρ_n to be decreasing with (increasing) Δt using ρ_n^- and ρ_n^+, but we use the asymptotic behaviour of ρ_n for this purpose.

Using relationship (A4) and the lower bound of y leads to,

$$|\bar{y}| \leqslant \tfrac{1}{2}. \tag{A5}$$

For large n and small Δt the asymptotic behaviour of ρ_n^- and ρ_n^+ shows that the spectral radius ρ_n behaves as,

$$\log \rho_n \sim -n \frac{\Delta t^2}{2}(1-2|\bar{y}|). \tag{A6}$$

It is then clear from (A6) that for large n, ρ_n is a decreasing function of (small) Δt. Furthermore, (A6) tells us that the spectral radius behaves, for large n, to accuracy $O(\Delta t^4)$ as;

$$\rho_n \sim \left| 1 - \frac{\Delta t^2}{2}(1-2|\bar{y}|)\right|^n. \tag{A7}$$

We can use (A7) to get an approximation to the upper and lower bounds on Δt^2. The condition for convergence of (A7), $|1-\frac{1}{2}\Delta t^2(1-2|\bar{y}|)| \leqslant 1$, gives an upper bound Δt_+^2 on Δt^2. Considering that $\bar{y}$ is close to zero, we have $\Delta t_+^2 = 4$ (actually a correct value is given by $4/(1-2|\bar{y}|)$. So, one has to expect that for $\Delta t \geqslant \Delta t_+$ divergence may occur. Now, for ρ_n to be decreasing with (increasing) Δt the condition, $0 \leqslant 1-\frac{1}{2}\Delta t^2(1-2|\bar{y}|) \leqslant 1$, is to be satisfied. This gives an upper limit, $\Delta t_-^2 = 2$, for ρ_n to decrease with Δt. In conclusion, to accuracy $O(\Delta t^4)$, the bounds give

$$\begin{cases} \Delta t \leqslant \sqrt{2} \Rightarrow \text{convergence faster for larger } \Delta t \\ \sqrt{2} \leqslant \Delta t \Rightarrow \text{convergence slower for larger } \Delta t \\ 2 < \Delta t \quad \Rightarrow \text{convergence may cease.} \end{cases} \tag{A8}$$

(B) *Resolvent expansion for the extended L86 system*

We suppose here that the L86 is weakly forced/dissipated as mentioned in eq (29), with small v_i's and f_i's. The Jacobian $G'(t)$, coefficient of Δt in the Taylor expansion of the resolvent, (12), is given by:

$$G'(t) = \begin{pmatrix} -v_1 & -x_3 & -x_2 & 0 & 0 \\ x_3 & -v_2 & x_1 & 0 & 0 \\ -x_2 & -x_1 & -v_3 & -b & 0 \\ 0 & 0 & b & -v_4 & -1 \\ 0 & 0 & 0 & 1 & -v_5 \end{pmatrix}, \tag{B1}$$

and the coefficient of $\frac{1}{2}\Delta t^2$ in the same expansion, $\dot{G}'(t) + G'^2(t)$, is given by:

$$\left(\begin{array}{cc} v_1^2 + -x_3^2 + x_2^2 & \begin{array}{c} v_{123}x_3 + 2x_1x_2 \\ + bx_4 - f_1 \end{array} \\ \begin{array}{c} -v_{123}x_3 - 2x_1x_2 \\ -bx_4 + f_3 \end{array} & v_2^2 - x_1^2 - x_3^2 \\ \begin{array}{c} v_{123}x_2 - 2x_1x_3 - f_2 \\ -bx_2 \end{array} & \begin{array}{c} v_{123} + 2x_2x_3 - f_1 \\ -bx_1 \end{array} \\ 0 & 0 \end{array} \right.$$

$$\left. \begin{array}{ccc} v_{123}x_2 - 2x_1x_3 - f_2 & bx_2 & 0 \\ v_{123}x_1 - 2x_2x_3 + f_1 & -bx_1 & 0 \\ v_3^2 + x_2^2 - x_1^2 - b^2 & bv_{34} & b \\ -v_{34}b & v_4^2 - b^2 - 1 & v_{45} \\ b & -v_{45} & v_5^2 - 1 \end{array} \right), \tag{B2}$$

where v_i $(i = 1, \ldots, 5)$ are the dissipation rates and f_i the components of the uniform forcing F in (28). Also in (B2), $v_{ijk} = v_i + v_j + v_k$, and $v_{ij} = v_i + v_j$. We will ignore the forcing now because in this case, weak forcing only slows the asymptotic system collapse to a rest state or to a periodic state (depending on the dissipation).

We focus on a case study with three variables updated, (x_1, x_3, x_4). In this case, the amplification matrix R_{uu} is given, to accuracy $O(\Delta t^3)$ by:

$$R_{uu}^n = R_{uu}(t_n, t_{n-1})$$

$$= \begin{pmatrix} \begin{array}{l} 1 - v_2\Delta t - \frac{\Delta t^2}{2}(x_1^2(t_n) \\ \quad + x_3^2(t_n) - v_2^2) \end{array} & 0 \\ 0 & \begin{array}{c} 1 - v_5\Delta t - (1 - v_5^2) \\ \times \frac{\Delta t^2}{2} \end{array} \end{pmatrix}. \tag{B3}$$

Notice from (B3), that in the Hamiltonian case, $(v_i = 0)$, the spectral radius of the product $A_n = \prod_{k=1}^n R_{uu}^n$ (see eq. (17)) controls the convergence to accuracy $O(\Delta t^4)$. Similar analysis to Appendix A shows that, for small Δt and large n, the logarithm of the convergence ρ_n is given to accuracy $O(\Delta t^4)$ by:

$$\log \rho_n = -n \frac{\Delta t^2}{2} \min(1, \overline{x_1^2 + x_3^2}), \tag{B4}$$

which satisfies (22) for $\Delta_1 t \leqslant \Delta_2 t \leqslant \Delta t_{\max}$. The expression $\overline{x_1^2 + x_3^2}$ represents the time average of $x_1^2 + x_3^2$. The turnover point can be calculated to $O(\Delta t^4)$ and is given by $\sqrt{2/\min(1, \overline{x_1^2 + x_3^2})}$. For the case of Fig. 12, this value is around 1.414, and the value of Δt_{max} observed from the convergence of the assimilation experiments corresponds roughly to $N = 280 - 290$, that is, $\Delta t_{max} \approx 1.45$.

Consider now convergence with weak dissipation and let us consider only v_2 nonzero, so that the amplification matrix contains one $O(\Delta t)$ term; $(1 - v_2 \Delta t)$. Even if Δt is small enough so that the $O(\Delta t)$ term dominates the second order term, the convergence, ρ_n, is still behaving, to accuracy $O(\Delta t^4)$ as:

$$\log \rho_n \sim -n \frac{\Delta t^2}{2},$$

because the spectral radius of R^n_{uu} depends on $1 - \frac{1}{2}\Delta t^2$ (and not $1 - v_2 \Delta t$) for small enough Δt. In this case increasing convergence with increasing Δt can still be expected. To allow the $O(\Delta t)$ term to play a role in convergence we suppose both v_2 and v_5 are nonzero, say $v_2 = v_5 = v$. The same analysis now shows that the convergence, to accuracy $O(\Delta t^3)$, and not $O(\Delta t^4)$ as before, behaves as:

$$\rho_n \sim \left(1 - v\Delta t - (\alpha - 3v^2)\frac{\Delta t^2}{2}\right)^n \tag{B5}$$

where $\alpha = \min(1, \overline{x_1^2 + x_3^2})$. For small Δt, and to accuracy $O(\Delta t^3)$, (B5) becomes;

$$\rho_n \sim 1 - \sigma v\left(1 - \frac{\sigma v}{2}\right) - \sigma\Delta t\left(\frac{\alpha}{2} - v^2\right), \tag{B6}$$

where σ plays the role of system time; $\sigma = n\Delta t$.

Different behaviours can now be expected from (B6) according to the relative sizes of v, α and Δt. First of all notice that the second term in (B6) depends only on system time and not on Δt. Therefore if $v \gg \Delta t$ such that this term is the dominant modification to 1, then the convergence will be almost independent of Δt. This is the result found by Williamson and Dickinson (1972), the physical interpretation of which is that there is data redundancy for frequent updating assimilation and convergence occurs at a constant rate. If $v \ll \Delta t$, such that the third term dominates the second term, and provided $v < \sqrt{\alpha/2}$, then the Hamiltonian type behaviour is recovered in which convergence improves for increasing Δt. This is the situation which has been demonstrated here experimentally. Note that even when the second term dominates the third, there may be small improvements in convergence rate with Δt provided $v < \sqrt{\alpha/2}$.

(C) *Resolvent approximation of some strongly dissipative systems*

In this appendix, we give an approximation of the resolvent of the Lorenz (1963) and Rössler (1976) systems to accuracy $O(\Delta t^2)$ only. The equations for the Lorenz (1963) are:

$$\begin{cases} \dot{x} = \sigma(-x + y) \\ \dot{y} = -xz + rz - y, \\ \dot{z} = xy - bz \end{cases} \tag{C1}$$

where $\sigma = 10$, $r = 28$ and $b = \frac{8}{3}$ are fixed as in the original paper. The Jacobian $F'(t)$, coefficient of Δt in (12), is given by:

$$F'(t) = \begin{pmatrix} -\sigma & \sigma & 0 \\ r - z & -1 & -x \\ y & x & -b \end{pmatrix}, \tag{C2}$$

and the resolvent to accuracy $O(\Delta t^2)$ is simply; $I_3 + \Delta t F'(t)$. When y is assimilated, the convergence ρ_n behaves (for small Δt), to accuracy $O(\Delta t^2)$, as:

$$\rho_n \sim (1 - b\Delta t)^n.$$

The equations for the Rössler (1976) system are:

$$\begin{cases} \dot{x} = -(y + z) \\ \dot{y} = x + \alpha y \\ \dot{z} = \alpha + z(x - \mu). \end{cases} \tag{C3}$$

The parameter μ controls the chaotic behaviour of the attractor through period doubling bifurcations, and $\alpha = \frac{1}{5}$. To ensure a chaotic attractor, μ is chosen to be greater than the limit, $\mu_\infty = 4.2$ of the Feigenbaum sequence (Hellman, 1984). The value $\mu = 5$ is taken here (see for example Crutchfield et al., 1980 and Lichtenberg and Liberman, 1983). The Jacobian, $F'(t)$, is then given by:

$$F'(t) = \begin{pmatrix} 0 & -1 & -1 \\ 1 & \alpha & 0 \\ z & 0 & x - \mu \end{pmatrix} \tag{C4}$$

which can be used to calculate the resolvent to accuracy $O(\Delta t^2)$ as in the Lorenz system.

REFERENCES

Bokhove, O. and T. G. Shepherd, 1996. On Hamiltonian balanced dynamics and the slowest invariant manifold. *J. Atmos. Sci.* **53**, 277–297.

Browning, G. L. and H.-O. Kreiss, 1996. Analysis of periodic updating for systems with multiple timescales. *J. Atmos. Sci.* **53**, 335–348.

Charney, J., M. Halem and R. Jastrow, 1969. Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.* **26**, 1160–1163.

Cooper, M. and K. Haines, 1996. Altimetric assimilation with water property conservation. *J. Geophys. Res.* **101**, 1059–1077.

Courtier, P., J. N. Th/'epaut and A. Hollingsworth, 1994. A strategy for operational implementation of 4D-var using an incremental approach. *Quart. J.R. Meteor. Soc.* **120**, 1367–1388.

Crutchfield, J., D. Farmer, N. Packard, R. Shaw, G. Jones and R. J. Donnelly, 1980. Power spectrum analysis of a dynamical system. *Phys. Lett.* **76A**, 1–4.

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

Gustavson, F. G., 1966. On constructing formal integrals of a Hamiltonian system near an equilibrium point. *Astron. J.* **71**, 670–686.

Hellman, R. H. G., 1984. Feigenbaum sequences in conservative and dissipative systems. In *Chaos and order in nature* (H. Haken, ed.), pp. 232–248. Springer-Verlag. Berlin Heidelberg New York.

Hénon, M. and C. Heiles, 1964. The applicability of the third integral of motion: Some numerical experiments. *Astron. J.* **69**, 73–79.

Holmes, P., 1982. Proof of non-integrability for the Hénon-Heiles Hamiltonian system near an exceptional integrable case. *Physica D* **5**, 335–347.

Horn, R. A. and C. R. Johnson, 1988. Matrix analysis. *Cambridge Universisty Press. Cambridge.*

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

Kalman, R. E, 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* (Trans. ASME) **D82**, 35–45.

Lawden, D. F., 1975. *Analytical methods of optimization.* Scottish Academic Press. Edinburgh.

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

Lewis, J. M. and J. C. Derber, 1985. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus* **37A**, 309–332.

Lichtenberg, A. J. and M. A. Lieberman, 1983. *Regular and stochastic motion.* Springer-Verlg. New York.

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

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

Lorenz, E. N. 1986. On the existence of a slow manifold. *J. Atmos. Sci.* **43**, 1547–1557.

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

Ott. E., 1994. *Chaos in dynamical systems.* Cambridge University Press. Cambridge.

Rössler, O. E., 1976. An equation for continuous chaos. *Phys. Lett.* **57A**, 397–398.

Sasaki, Y. K., 1958. An objective analysis based on the variational method. *J. Meteor. Soc. Japan* **36**, 738–742.

Shepherd, T. G. 1990. Symmetries laws and Hamiltonian structures in geophysical fluid dynamics. *Adv. in Geophys.* **32**, Academic Press, 287–338.

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

Talagrand, O., 1981b. On the mathematics of data assimilation. *Tellus* **33**, 321–339.

Williamson, D. and A. Kasahara, 1971. Adaptation of meteorological variables forced by updating. *J. Atmos. Sci.* **28**, 1313–1324.

Williamson, D. and R. E. Dickinson, 1972. Periodic updating of meteorological variables. *J. Atmos. Sci.* **29**, 190–193.