THE UNIVERSITY OF READING

Application of Variational Data Assimilation to the Lorenz Equations Using the Adjoint Method

by

M.A. Wlasak & N.K. Nichols

Numerical Analysis Report 4/98

DEPARTMENT OF MATHEMATICS

Application of Variational Data Assimilation to the Lorenz Equations Using the Adjoint Method

by

M.A. Wlasak & N.K. Nichols

Numerical Analysis Report 4/98

The University of Reading
P O Box 220
Reading RG6 6AX
Berkshire

Appears in Numerical Methods in Fluid Dynamics VI, ICFD, Oxford, 1998, pp 555-561.

Application of Variational Data Assimilation to the Lorenz Equations Using the Adjoint Method.

M. A. Wlasak N. K. Nichols
The University of Reading

1 Introduction

The aim of data assimilation is to combine measured data with mathematical models in order to obtain accurate estimates for the current and future states of a physical system. Variational data assimilation finds the particular solution to a set of model equations that gives the best least squares fit to a set of observations. Variational techniques, in which adjoints are used, are currently being implemented to treat very large scale systems occuring in numerical weather forecasting [4]. An important question that arises in data assimilation concerns the period of time over which the process remains accurate. The behaviour of systems that are sensitive to perturbations is notoriously difficult to estimate accurately over long time scales. In this study, an attempt is made to extend the time period over which data assimilation is successful, so that values from the model correctly match observations in a consistent manner.

2 Model equations

The assimilation process is applied to a discrete form of the third-order Lorenz equations [5], given by

$$\begin{aligned} \dot{x} &= a(y-x) \\ \dot{y} &= rx - y - xz \\ \dot{z} &= xy - bz. \end{aligned} \tag{2.1}$$

The Lorenz equations constitute a Fourier truncation of the flow equations governing thermal convection, where a is the Prandtl number, r is a normalized Rayleigh number, and b is a nondimensional wavenumber [3]. There is an inherent diversity in the nature of the solutions to these equations, dependent on the values of the parameters a, b, r. In the case where the parameters are set at 10, 8/3, 28, respectively, a deterministic chaotic system results. The qualitative behaviour of the system is then highly sensitive to the initial conditions. Three unstable stationary points exist, at the origin and at $(\pm 6\sqrt{2}, \pm 6\sqrt{2}, 27)$. At the origin, the linearized system has one positive and two negative real eigenvalues, making the equilibrium a saddle point with a single unstable mode. At the two non-zero equilibria, the linearized system has one negative eigenvalue and two complex eigenvalues with positive real parts, inducing unstable spiral modes at each point. The system is dissipative with all trajectories entering a bounded ellipsoid in $\Re^3$. Further properties of the Lorenz equations are described in [5].

Discrete approximations to the equations (2.1) are formed on a uniform grid using the second-order explicit Runge-Kutta scheme known as the modified Euler method [1]. The discrete model equations are defined by

$$\begin{aligned}
x_{i+1} &= f(x_i, y_i, z_i, \Delta t) \\
&\equiv x_i + \frac{1}{2}a\Delta t(2(y_i - x_i) - a\Delta t(y_i - x_i) + \Delta t(rx_i - y_i - x_i z_i)) \\
y_{i+1} &= g(x_i, y_i, z_i, \Delta t) \\
&\equiv y_i + \frac{1}{2}\Delta t(2(rx_i - y_i - x_i z_i) + \\
&\quad \Delta t(ra(y_i - x_i) - (rx_i - y_i - x_i z_i) - \\
&\quad x_i(x_i y_i - bz_i) - az_i(y_i - x_i) - a\Delta t(y_i - x_i))(x_i y_i - bz_i))) \qquad (2.2) \\
z_{i+1} &= h(x_i, y_i, z_i, \Delta t) \\
&\equiv z_i + \frac{1}{2}\Delta t(2(x_i y_i - bz_i) + \Delta t(ay_i(y_i - x_i) + x_i(rx_i - y_i - x_i z_i) \\
&\quad + \Delta t a(y_i - x_i)(rx_i - y_i - x_i z_i) - b(x_i y_i - bz_i))),
\end{aligned}$$

for $i = 0, 1, \ldots, N-1$, where x_i, y_i, z_i approximate the states of the continuous system at time $t_i = i\Delta t \in [0, T]$ and $\Delta t = T/N$.

This discrete system, although sharing many of the properties of the true continuous Lorenz equations, does not necessarily provide a solution to the continuous problem and no attempt is made here to do so. Both continuous and discrete systems are chaotic in nature and are sensitive to perturbations, and thus the discrete model provides a test case which is qualitatively similar to the continuous case.

In recent years discrete models of the Lorenz equations have been used popularly as test problems for the application of advanced data assimilation techniques [3],[2],[7],[6]. Like many systems in meteorology and oceanography, these models are dissipative and volume reducing, tending to a set of zero volume. Variational assimilation is applied to discrete Lorenz models in [3] and [2], and it is shown that the ability to track the chaotic trajectories of the system is limited to short times. The objective function that measures the least-square error between the observations and the model solutions is found to exhibit more and more secondary minima as the assimilation interval increases, making it difficult to find the optimal. In [6] the accuracy of the solution to the variational problem for the Lorenz system at the end of the assimilation period is studied as the length of the period is increased. The error between the assimilated solution and the observations is shown to saturate at a finite value, and hence the inclusion of past observations has a limited impact on predictability. In all of these studies the deviations of the model solutions from the observations are all equally weighted in the objective function. The aim here is to investigate the effect of time-varying weights on the length of the interval over which variational assimilation is accurate.

3 Data assimilation problem

A 4-D variational data assimilation scheme is applied. All the observations (data values) across the whole time interval are matched to the model solutions simultaneously. The scheme involves minimizing a quadratic objective function $\mathcal{J}$ subject to the model equations (2.2). The problem is stated as follows.

Problem 1 *Minimize the objective function*

$$\mathcal{J} = \sum_{i=0}^{i=N-1} d_i((\tilde{x}_i - x_i)^2 + (\tilde{y}_i - y_i)^2 + (\tilde{z}_i - z_i)^2)\Delta t \qquad (3.1)$$

subject to the constraints

$$\begin{aligned} x_{i+1} &= f(x_i, y_i, z_i, \Delta t) \\ y_{i+1} &= g(x_i, y_i, z_i, \Delta t) \\ z_{i+1} &= h(x_i, y_i, z_i, \Delta t), \end{aligned} \tag{3.2}$$

where $\tilde{x}_i, \tilde{y}_i, \tilde{z}_i$ *are given observation values,* x_i, y_i, z_i *are the model values and* d_i *are the respective weights at time* $t_i = i\Delta t$, *for* $i = 0, 1, \ldots, N-1$.

The constrained minimization problem can be converted into an unconstrained problem using the method of Lagrange multipliers. Since the initial values x_0, y_0, z_0 completely determine the solution to the model equations and are the only degrees of freedom available, the problem can be restated in the following form.

Problem 2 *Find the initial values* x_0, y_0, z_0 *and a set of parameters* $\{lx_i,\ ly_i,\ lz_i\}_{i=1}^N$ *to minimize the cost functional*

$$\begin{aligned} \mathcal{L} &= \mathcal{J} + \sum_{i=0}^{i=N-1} lx_{i+1}(f(x_i, y_i, z_i) - x_{i+1}) \\ &\quad + ly_{i+1}(g(x_i, y_i, z_i) - y_{i+1}) + lz_{i+1}(h(x_i, y_i, z_i) - z_{i+1}), \end{aligned} \tag{3.3}$$

subject to x_i, y_i, z_i *satisfying the model equations (2.2) for* $i = 0, \ldots, N-1$.

The solution to this problem can be interpreted as the set of initial values that provides a best least-squares fit of the model solutions to the given observations.

Necessary conditions for a solution to Problem 2 involve the corresponding adjoint equations of the system, given by

$$lx_N = 0, \qquad ly_N = 0, \qquad lz_N = 0, \tag{3.4}$$

and

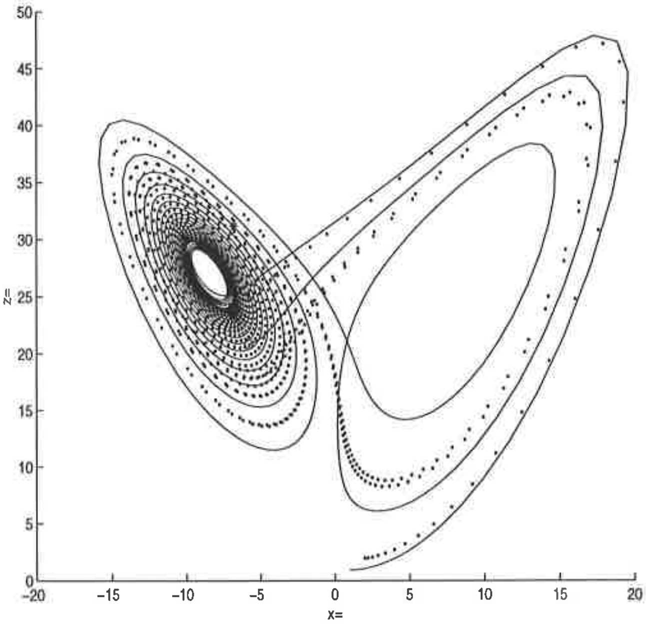
$$\begin{aligned} lx_i &= 2d_i\Delta t(\tilde{x}_i - x_i) + lx_{i+1}\left(\frac{\partial f}{\partial x_i}\right)_i + ly_{i+1}\left(\frac{\partial g}{\partial x_i}\right)_i + lz_{i+1}\left(\frac{\partial h}{\partial x_i}\right)_i \\ ly_i &= 2d_i\Delta t(\tilde{y}_i - y_i) + lx_{i+1}\left(\frac{\partial f}{\partial y_i}\right)_i + ly_{i+1}\left(\frac{\partial g}{\partial y_i}\right)_i + lz_{i+1}\left(\frac{\partial h}{\partial y_i}\right)_i \\ lz_i &= 2d_i\Delta t(\tilde{z}_i - z_i) + lx_{i+1}\left(\frac{\partial f}{\partial z_i}\right)_i + ly_{i+1}\left(\frac{\partial g}{\partial z_i}\right)_i + lz_{i+1}\left(\frac{\partial h}{\partial z_i}\right)_i, \end{aligned} \tag{3.5}$$

for $i = N-1, N-2, \ldots, 0$. The gradient of the objective functional $\mathcal{L}$ with respect to the initial data $(x_0, y_0, z_0)^T$ is then given by

$$\nabla\mathcal{L} = (lx_0, ly_0, lz_0).$$

For any specified set of intial values (x_0, y_0, z_0), by forward time-stepping the model equations (2.2) and then backward time-stepping the adjoint equations (3.5), the initial values (lx_0, ly_0, lz_0) of the adjoint equations can be found. These initial values must be equal to zero if the objective function is minimized. Otherwise they provide the local descent direction needed to find an improved estimate for the optimal initial values of the model system using a gradient minimization technique. The simplest gradient iteration method, the method of steepest descent, is applied here. The convergence criterion used to stop the iteration is given by the condition $\|(lx_0, ly_0, lz_0)^T\|_2 \leq 0.001$. (Here $\|\cdot\|_2$ denotes the L_2 - norm.)

Figure 1: Solutions over a long time interval: $0 \leq t \leq 16$



4 Application of Data Asssimilation

The main aim is to capture the general properties of the 'true' solution through the data assimilation process. With the Lorenz model, as the time interval is increased, the objective functional becomes increasingly sensitive to perturbations in the initial conditions [3] and inaccuracies in the best fit to the observed data accumulate until the assimilated solution takes qualitatively the wrong pattern.

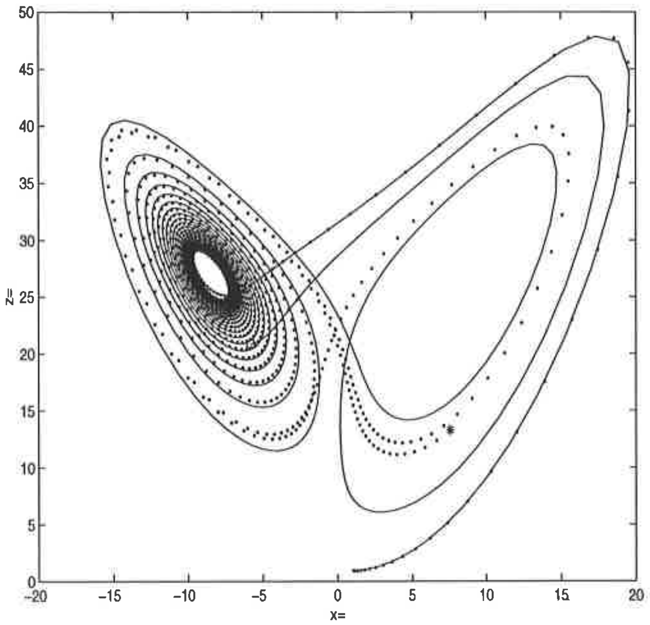
The sensitivity of the problem can be seen in Fig. 1 where the time period is set to $T = 16$ with $N = 1000$ equally spaced time intervals. The observations are generated by the model with the initial values (x_0, y_0, z_0) set to $(1, 1, 1)$. An estimate, or 'first guess' solution, is found from the model equations using the initial values $(2, 2, 2)$. In the figure the observations are denoted by a solid line; the 'first guess' solution is displayed as a dotted line. The final values of the two solutions at $T = 16$ are indicated, respectively, by a circle and a star. The figure shows the $x - z$ phase plane diagram of the solutions. The results are seen to diverge widely from each other.

Fig. 2 demonstrates the effect of data assimilation with constant weights $d_i = 1$, $i = 0, \ldots, N-1$, in the objective function (3.1). The observations are again represented by a solid line, and the solution to the data assimilation problem is shown by a dotted line. Qualitatively the match between the two solutions is better than in Fig. 1, especially at the start of the time period. The final value of the assimilated solution at the end time $T = 16$ (indicated by the star) is, however, still far away from the observed value (shown by the circle).

The choice of the weights d_i in the objective function $\mathcal{J}$ influences the overall dynamics of the assimilated solution and an appropriate choice may extend the time period over which accurate data assimilation is possible. A set of weights generated by a decaying exponential function with respect to time is now considered. The weights are given by

$$d_i = \frac{N+1}{\sum_{j=0}^{j=N} exp(-\gamma j \Delta t)} exp(-\gamma i \Delta t), \tag{4.1}$$

Figure 2: Application of data assimilation with constant weighting



for $i = 0, \ldots, N-1$, where γ is a specified constant. The motivation for this choice is that the error in the discretised Lorenz equations increases roughly exponentially with time. A decaying exponential weighting function compensates for an increase in error, ensuring that the more accurate values earlier in time have a stronger bearing on the final result. Such weights also guarantee that the final initial values chosen are closer to the observations at $t = 0$ and therefore are more likely to reflect the behaviour represented by the observations.

Obviously, the question of whether the use of time-varying weights is beneficial depends on the value given to the parameter γ. For the case where the weights are given by (4.1) with $\gamma = 1$, the solutions to the data assimilation problem over the time period $0 \leq t \leq T = 16$, with a timestep of 0.016, are shown in Fig. 3. The observations are again generated from the 'true' initial values $(1, 1, 1)$ and the minimization procedure is initiated with a 'first guess' solution obtained from the initial values $(2, 2, 2)$.

A comparison between Fig. 2 and Fig. 3 shows that the application of time-varying weights produces a better overall result than the use of constant weights. Over the time period the data assimilation process with constant weighting picks a solution that follows a different trajectory to that of the observations. The time-weighted solution is more accurate, but is diverging from the true trajectory at the end of the time interval. The values at the end of the time period weakly influence the overall data assimilation process, allowing for the aggregation of error. The insight given by these results suggests that a less extreme weighting distribution needs to be selected.

In Fig. 4, the solutions to the data assimilation problem over the time period $0 \leq t \leq 16$ are shown for the case where the weights are given by (4.1) with $\gamma = 0.5$. The time-step is kept at 0.016 and the observations are again generated from the 'true' initial value $(1, 1, 1)$. The same initial estimate of the solution is taken. When the new weighting is applied, the trajectories obtained by the data assimilation process match the observations almost perfectly.

This shows that the weighting must be balanced so as not only to ensure a more accurate initial condition, but also to ensure that the errors in the model solutions at the end of the time period do not dominate over the gains in accuracy obtained at earlier times.

Figure 3: Application of time varying weights with $\gamma = 1$

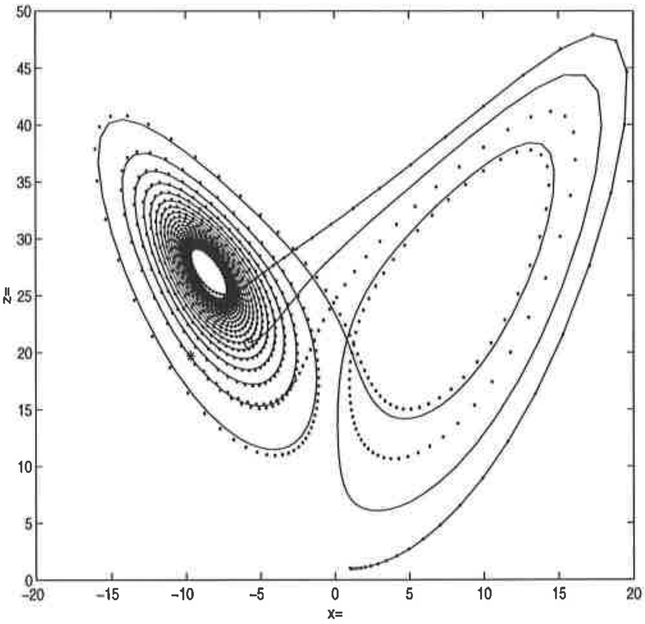
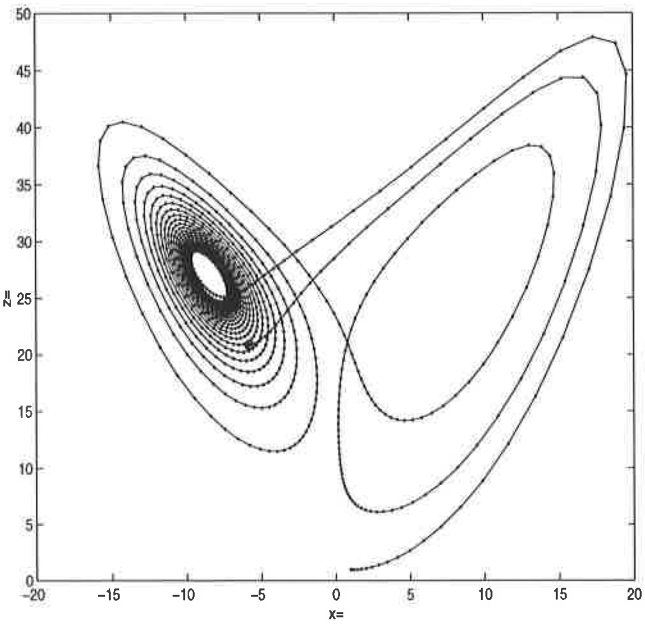


Figure 4: Application of time-varying weights with $\gamma = 0.5$



5 Conclusions

The results of the study show that a decaying set of time-varying weights can extend the time period over which data assimilation is successful and can improve the accuracy of the assimilated solution over the whole time period. For chaotic systems of the type arising in numerical weather prediction, the accuracy of the solution at the end of an assimilation period is critical, as this information is used to initialize the next forecast.

To show that time-varying weights can be used more generally, a more extensive investigation is needed. The effects of errors in the observations and the effects of larger data sampling intervals need to be examined. Ideally, a method for finding an *a priori* weight function that maximizes the extent of the accurate data assimilation period is also needed. These are topics of current research.

References

[1] S.D. Conte and C. de Boor. *Elementary Numerical Analysis.* McGraw-Hill, third edition, 1981.

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

[3] R.N. Miller M. Ghil and F. Gauthiez. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of Atmospheric Sciences*, 51(8):1037–1056, 1994.

[4] A.C. Lorenc. Analysis methods for numerical weather prediction. *Quart. J. Met. Soc.*, 112:1177–1194, 1986.

[5] C. Sparrow. *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors.* Springer-Verlag, 1982.

[6] C. Pires R. Vautard and O. Talagrand. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus*, 48A:96–121, 1996.

[7] M.A. Wlasak. Variational data assimilation: A study. Numerical Analysis Report 10, University of Reading, 1997.