Ensemble data assimilation with the Lorenz equations

Stefano Migliorini, University of Reading

June 2, 2010

1 The Lorenz model

The Lorenz (1963) model consists of three coupled nonlinear ordinary differential equations with chaotic dynamics:

$$\begin{aligned} \frac{dx}{dt} &= f_x(x, y, z, \sigma) = -\sigma(x - y) \\ \frac{dy}{dt} &= g_x(x, y, z, \rho) = \rho x - y - xz \\ \frac{dz}{dt} &= h_x(x, y, z, \beta) = xy - \beta z \end{aligned} \tag{1}$$

where $x = x(t)$, $y = y(t)$, $z = z(t)$ and σ, ρ, β are parameters, which in these experiments are chosen to have the values 10, 28 and 8/3 respectively.

The system is discretized using a fourth order Runge-Kutta method, which gives the following discrete equations:

$$\begin{aligned} x_{k+1} &= x_k + \Delta t(f_1 + 2f_2 + 2f_3 + f_4) \\ y_{k+1} &= y_k + \Delta t(g_1 + 2g_2 + 2g_3 + g_4) \\ z_{k+1} &= z_k + \Delta t(h_1 + 2h_2 + 2h_3 + h_4) \end{aligned} \tag{2}$$

where

$$\begin{aligned} f_1 &= f_x(x, y, z, \sigma) \\ g_1 &= f_y(x, y, z, \rho) \\ h_1 &= f_y(x, y, z, \beta) \\ f_2 &= f_x(x + \Delta t\frac{f_1}{2}, y + \Delta t\frac{g_1}{2}, z + \Delta t\frac{h_1}{2}, \sigma) \\ g_2 &= f_y(x + \Delta t\frac{f_1}{2}, y + \Delta t\frac{g_1}{2}, z + \Delta t\frac{h_1}{2}, \rho) \end{aligned}$$

$$
\begin{aligned}
h_2 &= f_z(x + \Delta t\frac{f_1}{2}, y + \Delta t\frac{g_1}{2}, z + \Delta t\frac{h_1}{2}, \beta) \\
f_3 &= f_x(x + \Delta t\frac{f_2}{2}, y + \Delta t\frac{g_2}{2}, z + \Delta t\frac{h_2}{2}, \sigma) \\
g_3 &= f_y(x + \Delta t\frac{f_2}{2}, y + \Delta t\frac{g_2}{2}, z + \Delta t\frac{h_2}{2}, \rho) \\
h_3 &= f_z(x + \Delta t\frac{f_2}{2}, y + \Delta t\frac{g_2}{2}, z + \Delta t\frac{h_2}{2}, \beta) \\
f_4 &= f_x(x + \Delta t\frac{f_3}{2}, y + \Delta t\frac{g_3}{2}, z + \Delta t\frac{h_3}{2}, \sigma) \\
g_4 &= f_y(x + \Delta t\frac{f_3}{2}, y + \Delta t\frac{g_3}{2}, z + \Delta t\frac{h_3}{2}, \rho) \\
h_4 &= f_z(x + \Delta t\frac{f_3}{2}, y + \Delta t\frac{g_3}{2}, z + \Delta t\frac{h_3}{2}, \beta)
\end{aligned} \tag{3}
$$

where Δt is the model time step and k is the time step index.

The time discretization procedure introduces an error to the solution of the Lorenz equations. An optional model error term, with given standard deviation, can be added to the discretized equations that can then be written as

$$
\begin{aligned}
x_{k+1} &= x_k + \Delta t(f_1 + 2f_2 + 2f_3 + f_4) + \sqrt{\Delta t}\eta_x \\
y_{k+1} &= y_k + \Delta t(g_1 + 2g_2 + 2g_3 + g_4) + \sqrt{\Delta t}\eta_y \\
z_{k+1} &= z_k + \Delta t(h_1 + 2h_2 + 2h_3 + h_4) + \sqrt{\Delta t}\eta_z
\end{aligned} \tag{4}
$$

where $\eta = (\eta_x, \eta_y, \eta_z)^T \sim N(\mathbf{0}, \mathbf{Q})$ is assumed to be a normally distributed random vector with zero mean and error covariance $\mathbf{Q}$. Here we use a diagonal $\mathbf{Q}$ with variance values for x, y and z given by 0.1491, 0.9048 and 0.9180, respectively (see Evensen, 2007).

2 The ensemble Kalman filter (EnKF)

In this program the square root algorithm for the EnKF analysis presented in Evensen (2004) is used. The analysis $\overline{\mathbf{A}}^a$, defined as the mean of $\mathbf{A}^a = (\mathbf{x}_1^a, \mathbf{x}_2^a, \cdots, \mathbf{x}_N^a)$, where $\mathbf{x}_i^a$ is the i-*th* analysis ensemble member, can be calculated as

$$
\overline{\mathbf{A}}^a = \overline{\mathbf{A}}^f + \mathbf{A}'\mathbf{S}^T\mathbf{C}^{-1}(\mathbf{y} - \mathbf{H}\overline{\mathbf{A}}^f) \tag{5}
$$

where $\overline{\mathbf{A}}^f$ is the mean of $\mathbf{A}^f = (\mathbf{x}_1^f, \mathbf{x}_2^f, \cdots, \mathbf{x}_N^f)$, being $\mathbf{x}_i^f$ the i-*th* forecast ensemble member with initial conditions given by the analysis ensemble members at the previous data assimilation cycle. The matrix $\mathbf{A}'$ is defined as equal to $\mathbf{A}^f - \overline{\mathbf{A}}^f$ and $\mathbf{S} = \mathbf{H}\mathbf{A}'$, where $\mathbf{H}$ is the observation operator that simulates observations from model fields, and $\mathbf{C} = \mathbf{S}\mathbf{S}^T + (N-1)\mathbf{R}$ where $\mathbf{R}$ is the observation error covariance and $\mathbf{y}$ is the observation vector. The analysis perturbation matrix is calculated as

$$\mathbf{A}^{a\prime} = \mathbf{A}'\mathbf{V}\sqrt{\mathbf{I} - \mathbf{\Sigma}^T\mathbf{\Sigma}}\mathbf{V}^T \tag{6}$$

where $\mathbf{C}^{-1} = \mathbf{Z}\mathbf{\Lambda}^{-1}\mathbf{Z}^T$ and $\mathbf{X} = \mathbf{\Lambda}^{-1/2}\mathbf{Z}^T\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

3 The program

The programs are written in IDL language and consist of a main program (lorenz_enkf.pro) and a related subroutine (lorenz.pro). The main program has a set of optional input arguments, described below. These arguments have default values so that it is possible to run the program by simply typing:

idl

IDL> lorenz_enkf.pro

provided both lorenz_enkf.pro and lorenz.pro are in a directory that is included in the IDL path environment variable. If this is not the case, the programs need to be compiled first, by entering

.r path/lorenz_enkf.pro

.r path/lorenz.pro

where path is the path to the directory where the programs are stored. The integration time step is equal to 0.01 and the assimilation window is between time=0 and time=4. Between time=4 and time=8 a forecast ensemble is integrated forward, with no further assimilation.

The main program can be run with an optional argument *arg* or *flag* (which can be true or false) by entering

lorenz_enkf, *arg* = *argval*, */flag*

where *argval* is an appropriate argument value. The arguments and flags are the following:

- nrens=nrens

  This defines the number of ensemble members (default 100).

- nmeas=nmeas

  Total number of measurements (default 80)

- ini_var=ini_var

  Initial state error variance, at the moment is the same for x, y and z (default 2.0)

- meas_var=meas_var

  Observation error variance, at the moment is the same for x, y and z (default 2.0)

- perf_obs=perf_obs

  Flag to define perfect observations (i.e., $\mathbf{y} = \mathbf{x} = (x, y, z)^T$). Note that for numerical consistency the observation error variances are different from zero and set equal to one tenth of the initial state error variances. Default 0.

- test=test

  Flag to have the same observation values for a given total number of observations and a fixed initial conditions. Useful to compare results with different argument values (e.g., for different number of ensemble members). It creates a file with the RMSE between the analysis and the reference solution for the full state vector and with the observation mask values, for each time step. Default 0.

- cov_flag=cov_flag

  Flag to return an estimate of the state vector error covariance (see next argument). Default 0.

- covs=covs

  Returns an estimate of the state vector error covariance, when cov_flag = 1. Default 0.

- ih=ih

  Returns the observation mask, a vector with an element for each time step. The element is 0 when no observation is present and is 1 vice versa.

- ref=ref

  Returns the reference solution (i.e., the "truth").

- moderr_flag=moderr_

  Flag to consider model error (default 0).

- print_flag=print_

  Flag to produce postscript plots (default 0).

- path=path

  Path where to store the plots or ana_err.dat, when print_flag or test flag are 1, respectively.

4 Suggested exercises

1. When test=0 (default), the initial conditions are generated by taking a random sample from a normal distribution with variance equal to ini_var. How sensitive is the estimate on the value of ini_var?

2. Try to experiment with different choices for the total number of observations and their accuracy. Are the results independent of ini_var?

3. How sensitive are the results on the number of ensemble members used for representing the evoultion of the state?

4. Check the effect of including a model error term on the accuracy of the solution.

References

[1] Lorenz , E., N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20, 130-141.

[2] Evensen, G., 2004: Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dynamics*, 54, 539-560.

[3] Evensen, G., 2007: Data Assimilation: The Ensemble Kalman Filter, Springer.