

# Data Assimilation Demonstration with the Double Pendulum

Ross Bannister, University of Reading, National Centre for Earth Observation

r.n.bannister@reading.ac.uk

Version 1.1, June 8th 2010

## What is variational data assimilation?

Data assimilation is a tool used to estimate the initial conditions of a model using the available (but imperfect) observations of the system being modelled. Data assimilation is useful for weather forecasting, which relies on the availability of a wide range of recent measurements of the global weather. Having good quality initial conditions enables forecasters to launch weather prediction models equipped with the best possible information.

Variational data assimilation (or VAR) is a means

of estimating the initial conditions by optimizing the fit between real observations and predicted observations, where the optimization is done in a variational fashion. In the case of 3D-VAR (three spatial dimensions), the evolving state of the system is ignored while the data assimilation is performed, and in the case of 4D-VAR (three spatial dimensions and time) [1], the evolving state of the system is taken into account. 4D-VAR is the data assimilation method of choice in most operational weather prediction centres [2][3].

## What is a double pendulum?

Numerical weather prediction models are complicated, cumbersome, and require extensive resources to work. To demonstrate the technique of 4D-VAR, we use a very simple physical system - the double pendulum. Unlike a weather prediction model, which could contain over a million degrees of freedom, the double pendulum is described by four variables. These are two angles,  $\theta_1$  and  $\theta_2$  (see Fig. 1) and their rates of change,  $\dot{\theta}_1$  and  $\dot{\theta}_2$ . The prognostic equations for these variables, and how they are integrated in time are shown elsewhere [4]. When a double pendulum is set in motion it swings about the two axes and exhibits complicated motion [5].

The double pendulum equations are non-linear and show sensitive dependence to their initial conditions. This means that the data assimilation

must do a good job of estimating the initial conditions in order to make an accurate forecast.

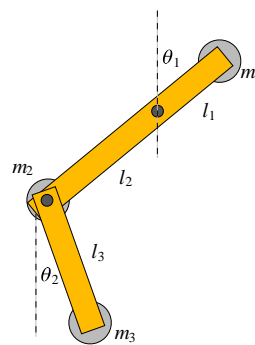


Figure 1: The double pendulum.

Even though the double pendulum works in a plane (ie in 2D), the VAR method used here takes into account the time dimension and so will still be called 4D-VAR, although it should really be called (2+1)D-VAR.

## Elements of the demonstration package

The following are essential files needed to run the double pendulum data assimilation package.

- The Fortran 77 source code (`Assim.f`).
- A configuration file that is read into the program to tell it how to run (`Assim.conf`).

The program may be run with the above files only, but is easier to run from within the MatLab

front-end which calls the program and plots the results. The following additional files allow this.

- Three configuration files (one for each program run mode) (`Nature.conf`, `Insert.conf`, `Var.conf`).
- Three MatLab scripts (`Nature.m`, `Insert.m`, `Var.m`).

This documentation assumes that the program is to be run inside MatLab.

## Compilation of the code

Compilation of the `Assim.f` Fortran source code to produce a file executable by the computer requires a Fortran compiler. How this is done

depends upon the system. E.g. with the `studio12` compiler on UNIX

```
f77 Assim.f -o Assim.out -xtypemap=real:64
```

## Using the demo 1: Running the double pendulum model and generating simulated observations

**Configuration file:** `Nature.conf`

**MatLab script:** `Nature.m`

To do data assimilation, observations of the double pendulum are required. In the absence of observations of a real pendulum, simulated observations are made. A set of initial conditions are chosen, and the model is integrated in time for a chosen period (this is called the 'truth' or 'nature' run). From this run, values of  $\theta_1$  and  $\theta_2$  are stored at selected times, and observation error is added to give a set of simulated observations.

To generate the nature run and simulated observations, edit `Nature.conf`. The following entries are important.

- Status (set to `MakeObs`).
- The lengths of the pendulum arms and the masses attached to them (Fig. 1).
- The acceleration due to gravity.
- The integration scheme's time-step.
- The length of the integration.

## Using the demo 2: Naive data assimilation by temporal insertion of observations

**Configuration file:** `Insert.conf`

**MatLab script:** `Insert.m`

This is an inferior method of assimilating the observations included to show how important VAR is (see next section). In this mode, an integration of the double pendulum is made starting from specified initial conditions (see below). At each stage of the integration when there is an observation, its value simply overwrites the model value at that time. No account is taken of the observation errors and the prognosis will not be dynamically consistent.

To run in this mode, edit `Insert.conf`. The following entries are important.

- Status (set to `Insert`).
- The lengths of the pendulum arms and the masses attached to them (as in `Nature.conf`).
- The acceleration due to gravity (as in `Nature.conf`).
- The integration scheme's time-step (as in `Nature.conf`).
- The length of the integration (as in `Nature.conf`).
- The guess at the initial conditions,  $\theta_1$ ,  $\theta_2$ ,  $\dot{\theta}_1$ ,  $\dot{\theta}_2$  (these should be similar, but different to those in `Nature.conf`).
- The size of the observation window

- The initial conditions,  $\theta_1$ ,  $\theta_2$ ,  $\dot{\theta}_1$ ,  $\dot{\theta}_2$ .
- The interval between which observations are simulated.
- The observation error (standard deviation) for  $\theta_1$  and  $\theta_2$  (set to zero for no observations made of  $\theta_1$  and  $\theta_2$ ).
- The observation error (standard deviation) for  $\dot{\theta}_1$  and  $\dot{\theta}_2$  (set to zero for no observations made of  $\dot{\theta}_1$  and  $\dot{\theta}_2$ ). It is usual for this parameter to be set to zero because observations of the rates of change of angles are difficult to make in a real system.

An example `Nature.conf` file is reproduced in the appendix.

Run the MatLab script `Nature.m` (this does not need to be edited). The program outputs the nature run in file `Assim.Nature` and the observations in file `Observations`. Four plots will appear showing  $\theta_1$ ,  $\theta_2$ ,  $\dot{\theta}_1$ ,  $\dot{\theta}_2$  and the observations (with error bars).

(the data assimilation is done in time segments, as in the data assimilation cycle).

- Lines of observational data.

Each line of observational data should have the following format.

value, type (1 is  $\theta_1$ , 2 is  $\theta_2$ , 3 is  $\dot{\theta}_1$ , 4 is  $\dot{\theta}_2$ ), error, time

The observation error is read-in but not used with this method of assimilation. The last observational data line must read

000.00, 0, 000.00, 000.00

which just marks the end of the list. An example `Insert.conf` file is reproduced in the appendix.

Run the MatLab script `Insert.m` (this does not need to be edited). The program outputs the observation-influenced integration in file `Assim.ins`. Four plots will appear showing  $\theta_1$ ,  $\theta_2$ ,  $\dot{\theta}_1$ ,  $\dot{\theta}_2$  (for nature and insertion runs) and the observations (with error bars).

Things to try.

- Run the insertion code with the imperfect observations output from the nature run (this may be done by copy and paste from the `Observations` file which has the same format as above; the observations in `Insert.conf` may be edited. Try this when the observations made of the nature run are

very accurate (e.g. observation error 5 degrees) and when they are inaccurate (e.g. 50 degrees).

- Try assimilating fewer observations

(e.g. delete the observations for the second half of the period).

- Introduce some outlying observations and see what happens.

### Using the demo 3: Variational data assimilation

**Configuration file:** `Var.conf`

**MatLab script:** `Var.m`

VAR is a method of assimilating the observations and is superior to the insertion method (previous section). In this mode, a model trajectory is sought which best fits the observations and a background state. The assimilation is performed in cycles, meaning that the background state for the 'next' cycle is a forecast of the analysis from the 'last' cycle. The initial background state needs to be specified. This method of data assimilation gives tries to obtain a dynamically consistent trajectory and the observation errors are taken into account.

To run in this mode, edit `Var.conf`. The following entries are important.

- Status (set to `Analyse`).
- The lengths of the pendulum arms and the masses attached to them (as in `Nature.conf`).
- The acceleration due to gravity (as in `Nature.conf`).
- The integration scheme's time-step (as in `Nature.conf`).
- The length of the integration (as in `Nature.conf`).
- The initial conditions,  $\theta_1$ ,  $\theta_2$ ,  $\dot{\theta}_1$ ,  $\dot{\theta}_2$  (these should be similar, but different to those in `Nature.conf`). This is the initial background state.
- The size of the observation window (the data assimilation is done in time segments, as in the data assimilation cycle).
- Lines of observational data.

Each line of observational data should have the following format.

value, type (1 is  $\theta_1$ , 2 is  $\theta_2$ , 3 is  $\dot{\theta}_1$ , 4 is  $\dot{\theta}_2$ ), error, time

The last observational data line must read

000.00, 0, 000.00, 000.00

which just marks the end of the list. An example `Var.conf` file is reproduced in the appendix.

Run the MatLab script `Var.m` (this does not need to be edited). The program outputs the analysis trajectory in file `Assim.anal` and the free forecast (unaffected by observations) from the initial background state in file `Assim.fore`. Two sets, each of four plots will appear showing  $\theta_1$ ,  $\theta_2$ ,  $\dot{\theta}_1$ ,  $\dot{\theta}_2$  (set 1 for nature and VAR runs and set 2 for the nature and background forecasts). On each plot the observations are also plotted (with error bars).

Things to try.

- Run the VAR code with the imperfect observations output from the nature run (this may be done by copy and paste from the `Observations` file which has the same format as above; the observations in `Var.conf` may be edited). Try this when the observations made of the nature run are very accurate (e.g. observation error 5 degrees) and when they are inaccurate (e.g. 50 degrees).
- Try assimilating fewer observations (e.g. delete the observations for the second half of the period).
- Introduce some outlying observations and see what happens when large and small errors are assigned to the observation.

### Appendix - example configuration files

**Example configuration file:** `Nature.conf`

```
=====
4dVar (Double Pendulum) Input File
=====
==== General Parameters =====
--Status ('MakeObs'='invent obs, 'Analyse'='4dVar, 'Insert'='simple insrt.)
MakeObs
--Dimensions of pendulum (l1,l2,l3, metres)
0.1, 0.2, 0.15
--Masses of pendulum nodes (m1,m2,m3, kg)
0.1, 0.1, 0.1
--Acceleration due to gravity (m/s/s)
10.0
```

```

--Time step (seconds)
0.001
--Length of integration (time steps)
5000
--Model state (initial guess(Analyse) or initial conditions(MakeObs))
--angle 1,angle 2, roc angle 1, roc angle 2 (deg and deg/s)
160.0, 100.0, -350.0, 100.0
=====
==== MakeObs Parameters =====
--Frequency of observation output (used only for MakeObs)
50
--Noise for angles (deg, used only for MakeObs)
5.0
--Noise for roc angles (deg/s used only for MakeObs)
0.0
=====
==== Analyse/Insert Parameters =====
--Size of each obs window (time steps used only for Analyse/Insert)
500
--Observations (used only for Analyse/Insert)
--Value, type (1,2,3,4), error, time (sec)
-----
000.00, 0, 000.00, 000.00
-----
=====

```

### Example configuration file: Insert.conf

```

=====
4dVar (Double Pendulum) Input File
=====
==== General Parameters =====
--Status ('MakeObs'=invent obs, 'Analyse'=4dVar, 'Insert'=simple insrt.)
Insert
--Dimensions of pendulum (l1,l2,l3, metres)
0.1, 0.2, 0.15
--Masses of pendulum nodes (m1,m2,m3, kg)
0.1, 0.1, 0.1
--Acceleration due to gravity (m/s/s)
10.0
--Time step (seconds)
0.001
--Length of integration (time steps)
5000
--Model state (initial guess(Analyse) or initial conditions(MakeObs))
--angle 1,angle 2, roc angle 1, roc angle 2 (deg and deg/s)
140.0, 130.0, -300.0, 140.0
=====
==== MakeObs Parameters =====
--Frequency of observation output (used only for MakeObs)
50
--Noise for angles (deg, used only for MakeObs)
50.0
--Noise for roc angles (deg/s used only for MakeObs)
0.0
=====
==== Analyse/Insert Parameters =====
--Size of each observation window (time steps)
500
--Observations (used only for Analyse)
--Value, type (1,2,3,4), error, time (sec)
-----
144.4436  1  5.0000  0.049000
107.5278  2  5.0000  0.049000
123.3436  1  5.0000  0.099000
109.0008  2  5.0000  0.099000
 91.2156  1  5.0000  0.149000
111.2956  2  5.0000  0.149000
 174.3908  1 100.0000  0.199000
 96.1762  2  5.0000  0.199000
149.0258  1 100.0000  0.249000
 71.0156  2  5.0000  0.249000

```

```

130.3854  1  100.0000  0.299000
27.1486  2  5.0000  0.299000
109.9339  1  100.0000  0.349000
-24.6323  2  5.0000  0.349000
-25.5732  1  5.0000  0.399000
-39.3050  2  5.0000  0.399000
-65.5397  1  5.0000  0.449000
-45.8575  2  5.0000  0.449000
-94.5846  1  5.0000  0.499000
-57.7381  2  5.0000  0.499000
-113.2558 1  5.0000  0.549000
-78.8796  2  5.0000  0.549000
-134.9893 1  5.0000  0.599000
-104.8157 2  5.0000  0.599000
-145.2104 1  5.0000  0.649000
-119.3656 2  5.0000  0.649000
-150.5228 1  5.0000  0.699000
-133.1293 2  5.0000  0.699000

```

etc.

```

7.7097  1  5.0000  1.999000
176.9603 2  5.0000  1.999000
000.00, 0, 000.00, 000.00

```

### Example configuration file: var.conf

```

=====
4dVar (Double Pendulum) Input File
=====
==== General Parameters =====
--Status ('MakeObs'=invent obs, 'Analyse'=4dVar, 'Insert'=simple insrt.)
Analyse
--Dimensions of pendulum (l1,l2,l3, metres)
0.1, 0.2, 0.15
--Masses of pendulum nodes (m1,m2,m3, kg)
0.1, 0.1, 0.1
--Acceleration due to gravity (m/s/s)
10.0
--Time step (seconds)
0.001
--Length of integration (time steps)
5000
--Model state (initial guess(Analyse) or initial conditions(MakeObs))
--angle 1,angle 2, roc angle 1, roc angle 2 (deg and deg/s)
140.0, 130.0, -300.0, 140.0
=====
==== MakeObs Parameters =====
--Frequency of observation output (used only for MakeObs)
50
--Noise for angles (deg, used only for MakeObs)
50.0
--Noise for roc angles (deg/s used only for MakeObs)
0.0
=====
==== Analyse/Insert Parameters =====
--Size of each observation window (time steps)
500
--Observations (used only for Analyse)
--Value, type (1,2,3,4), error, time (sec)
-----
144.4436  1  5.0000  0.049000
107.5278  2  5.0000  0.049000
123.3436  1  5.0000  0.099000
109.0008  2  5.0000  0.099000
91.2156  1  5.0000  0.149000
111.2956  2  5.0000  0.149000
174.3908  1  100.0000  0.199000
96.1762  2  5.0000  0.199000
149.0258  1  100.0000  0.249000

```

71.0156	2	5.0000	0.249000
130.3854	1	100.0000	0.299000
27.1486	2	5.0000	0.299000
109.9339	1	100.0000	0.349000
-24.6323	2	5.0000	0.349000
-25.5732	1	5.0000	0.399000
-39.3050	2	5.0000	0.399000
-65.5397	1	5.0000	0.449000
-45.8575	2	5.0000	0.449000
-94.5846	1	5.0000	0.499000
-57.7381	2	5.0000	0.499000
-113.2558	1	5.0000	0.549000
-78.8796	2	5.0000	0.549000
-134.9893	1	5.0000	0.599000
-104.8157	2	5.0000	0.599000
-145.2104	1	5.0000	0.649000
-119.3656	2	5.0000	0.649000
-150.5228	1	5.0000	0.699000
-133.1293	2	5.0000	0.699000

etc.

7.7097	1	5.0000	1.999000
176.9603	2	5.0000	1.999000
000.00, 0,		000.00, 000.00	

-----  
 =====

## References

- [1] Bannister R.N., 2001/2007 Elementary 4D-VAR, DARC Technical Report No. 2, [www.met.rdg.ac.uk/~ross/Documents/Var4d.pdf](http://www.met.rdg.ac.uk/~ross/Documents/Var4d.pdf).
- [2] Rabier F., 2005, Overview of global data assimilation developments in numerical weather prediction centres, Q. J. R. Meteorol. Soc. 131, 3215-3233.
- [3] Rawlins F., Ballard S.P., Bovis K.J., Clayton A.M., Li D., Inverarity G.W., Lorenc A.C., Payne T.J., 2007, The Met Office global four-dimensional variational data assimilation scheme, Q. J. R. Meteorol. Soc. 133, 347-362.
- [4] Bannister R.N., The double pendulum (modified), [www.met.rdg.ac.uk/~ross/Physics/DP2.html](http://www.met.rdg.ac.uk/~ross/Physics/DP2.html).
- [5] Bannister R.N., [www.met.rdg.ac.uk/~ross/Physics/DPdemo.avi](http://www.met.rdg.ac.uk/~ross/Physics/DPdemo.avi).