Error covariances in a simple 1D assimilation system

Data Assimilation Research Centre (DARC)
University of Reading, UK*

Introduction

The aim of this practical is to explore how different settings of the statistical parameters effect the final analysis in a simple 1D data assimilation system. The matlab routines 'analysis_2obs', 'analysis_mobs', and 'analysis_sim' show the results of assimilating a small set of idealised (direct) observations, given a 1D background field and assumed background and observation error statistics.

In each case, information from the observations, $\mathbf{y}$, is combined with information from a *background* (or first guess) field, $\mathbf{x^b}$, according to the BLUE (Best Linear Unbiased Estimator) formula

$$\mathbf{x^a} = \mathbf{x^b} + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x_b})\,, \tag{1}$$

where

$$\mathbf{K} = \mathbf{B}\mathbf{H^T}\left(\mathbf{H}\mathbf{B}\mathbf{H^T} + \mathbf{R}\right)^{-1}\,, \tag{2}$$

is known as the *gain* matrix. The updated state, $\mathbf{x^a}$, is called the *analysis*, $\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariance matrices, respectively, and $\mathbf{H}$ is the observation operator (in these examples, it is just a simple linear interpolation).

The elements of the background error covariance matrix $\mathbf{B}$ are given by

$$\mathbf{B}_{ij} = \sigma^2 \rho_{ij}\,, \tag{3}$$

where σ is the error standard deviation. The ρ_{ij} are given by a second-order autoregressive (SOAR) correlation function of the distance r_{ij} between the grid-points i and j

$$\rho_{ij} = \left(1 + \frac{r_{ij}}{L}\right)\exp\left(\frac{-r_{ij}}{L}\right), \tag{4}$$

where L is the correlation length scale. A similar expression is used for the elements of $\mathbf{R}$, where the observation error correlations (if any) are assumed to depend on the distance between observations.

Running the code

Each routine is run by typing its name at the matlab command line. If you answer *'Yes'* to the question *'Use default parameter values?'*, the program will run with its default values. If you answer *'No'*, you will be asked various questions, such as where you want to put the observations, and the error characteristics of the observations and background field. Note that if you want to change an array, you will need to enter the correct number of new values.

The routine will plot a graph showing the observations, background and analysis, which you will be given the option of saving as a '.fig' file. Finally, a summary of the selected analysis parameters is printed to the screen.

*based on original notes by R. Swinbank & R. Petrie, DARC, UK

Exercise 1: analysis_2obs

This exercise uses equation (1) to calculate an analysis on the domain $x \in [0, 10]$ given two observations. The observations must take values between -2 and +2 but can be placed anywhere in the domain. The background field is assumed to be zero everywhere.

A list of inputs into the program and their default values is given in Table 1 below.

variable	**description**	**default value**
y1	value of observation 1	1.0
x1	x position of observation 1	4.0
y2	value of observation 2	-1.0
x2	x position of observation 2	6.0
sigo	observation error standard deviation, σ_o	1.0
Lo	observation error correlation length scale	0.0
sigf	background error standard deviation, σ_f	1.0
Lf	background error correlation length scale	2.0

Table 1: default parameter values for exercise 1

Figure 1 shows the output when the default parameter values in Table 1 are used. The dotted lines are the analyses that would have been produced if just one of the observations had been available and the dashed line is the analysis that is produced when both observations are used.

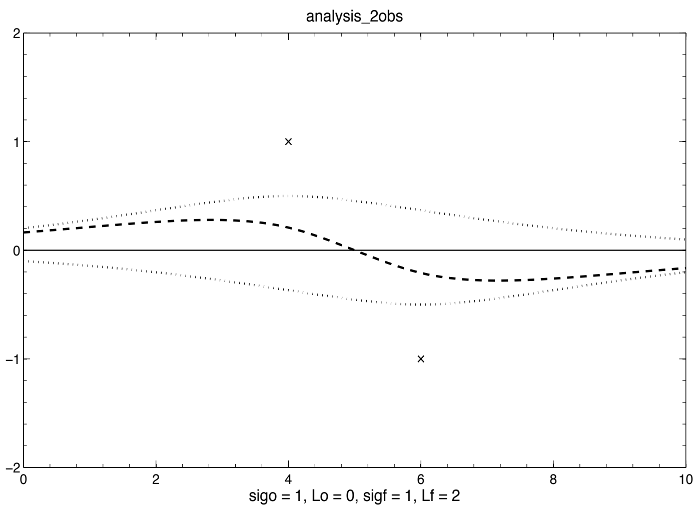

Figure 1: analysis_2obs with default parameters from Table 1.

Try to reproduce this figure by calling the script *'analysis_2obs'* from the matlab command line and selecting *'yes'* when asked if you want to use the default parameter values.

Now explore what happens if you change the inputs:

- what happens if you change the observation values, locations and assumed error standard deviation?
- what happens if you vary the background error correlation length? (e.g. from 0 to 10)
- what happens if you vary the background error standard deviations? try both increasing and decreasing.

The effect of the observation error covariances is harder to interpret:

- what happens if you assume the observation errors are highly correlated (large L), the observations are located close together and agree/ disagree?
- what happens if they located are far apart?

Try to explain the results you find.

Exercise 2: analysis_mobs

This exercise calculates an analysis given two sets of observations, A and B, with a maximum of seven in each set. As in the previous exercise, the observations should take values between -2 and +2 and lie within the domain $x \in [0, 10]$. The default set up is as follows:

SET A: 3 observations

position	4.0	5.0	6.0
value	1.0	1.0	1.0

observation error standard deviation, sigoa = 2.0
observation error length scale, Loa = 0.0

SET B: 3 observations

position	4.0	5.0	6.0
value	-1.0	-1.0	-1.0

observation error standard deviation, sigob = 1.0
observation error length scale, Lob = 2.0

The background field is assumed to be zero everywhere and has default error standard deviation, sigf = 1.0, and error lengthscale, Lf = 1.0.

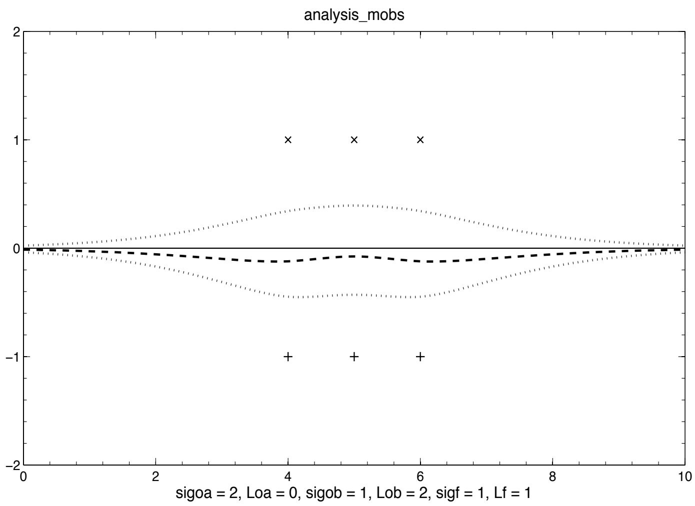

Figure 2: output from *analysis_mobs* with default parameters. The dotted lines are the analyses that are produced if only one set of observations is used, the dashed line is the analysis that is produced when both sets of observations are used.

Again, explore what happens if you change the inputs. For example,

- examine the effect of varying the number of observations, their values, error standard deviation, and spatial density.
- examine the effect of varying the observation error correlation lengthscales.

  how do these modifications relate to each other?

- try using one set of observations to represent sparse but accurate observations and the other set to represent dense but inaccurate observations.

Try to explain the results you find.

Exercise 3: analysis_sim

In this exercise we compute the analysis using a background field and observations generated from a given *'truth'* field.

You are given 3 options for defining the truth:

1. zero everywhere
2. pseudo step function
3. sum of several sinusoidal curves

The background field is generated either by adding random perturbations to the truth (according to the specified background error characteristics), or by shifting the truth field along the x-axis. The default background parameters are: error standard deviation, sigf = 2.0, and lengthscale, Lf = 2.0.

The observations are generated by adding random perturbations to the truth using the specified observation error statistics. There can be up to 11 observations, with horizontal (x axis) positions between 0 and 10. The default is 5 observations at positions [3.0, 4.0, 5.0, 6.0, 7.0] with observation error standard deviation, sigo = 1.0, and error lengthscale, Lo = 0.0 (uncorrelated).

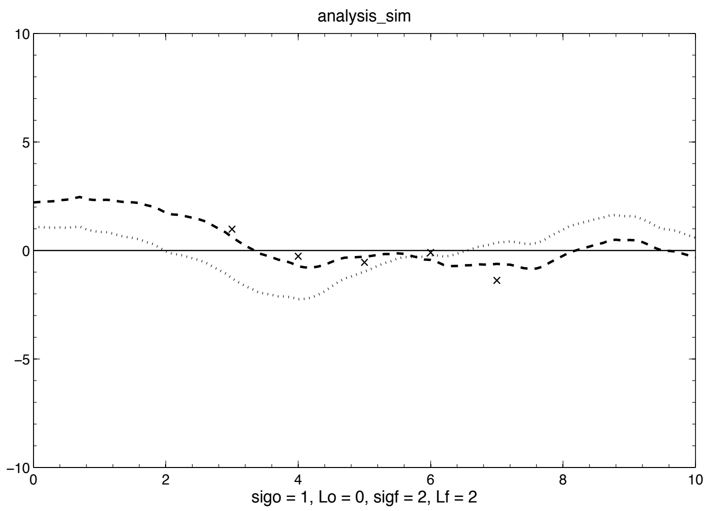

Figure 3: output from *analysis_sim* with truth equal to zero everywhere, background = truth + random perturbation, and default background and observation parameters. The observations are given by crosses, the dotted line is the background field and the dashed line is the analysis.

Once again, explore what happens if you change the inputs.

- try experimenting with different truth functions.
- try varying the number of observations, their errors and lengthscales with different background settings.
- an interesting exercise is to use the ‘step function’ form of the truth field with the background field equal to the truth, but shifted by a couple of units; which parameter settings are best for getting the analysis to move the ‘step’ back to the correct location?

Further exercises

The matlab routines used in these exercises are quite basic. Feel free to take copies and amend the code to enable you to carry out your own explorations. For example, you could easily amend *analysis_sim* to explore what happens if the assumed observation and background errors are different from their true values.