DATA ASSIMILATION TUTORIAL I
NOTES ON "TOY" ANALYSIS PROCEDURES

WRITTEN BY

R. SWINBANK MAY 2002

MODIFIED BY

R. PETRIE JUNE 2012

Introduction

The programs analysis_2obs, analysis_mobs and analysis_sim are a set of IDL/PV-wave procedures to show the results of analysing a small set of idealised observations, given a (one-dimensional) background field which is initially set to just be zero, and assumed (background and observation) error covariances. The aim is to allow you to explore the effect of varying the statistical parameters of simple data assimilation systems.

In each case, information from the observations ($\mathbf{y}_o$) is combined with information from a background field ($\mathbf{x}_b$) to produce an analysis field ($\mathbf{x}_a$) according to the standard OI (Optimal Interpolation) expression:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{B}\mathbf{H}^T \left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R} \right)^{-1} (\mathbf{y}_o - \mathbf{H}\mathbf{x}_b.) \tag{1}$$

$\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariances, respectively, and $\mathbf{H}$ is the observation operator (in these examples, just a simple interpolation). (Note that these IDL procedures explicitly invert the matrix, even though that is not usually the most efficient way of solving the equation.)

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

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

where σ the magnitude of background error. μ_{ij} is a second-order autoregressive correlation function of the distance r_{ij} between the grid-points i and j (by default):

$$\mu_{ij} = \left(\frac{1 + r_{ij}}{L} \right) \exp \left(\frac{-r_{ij}}{L} \right) \tag{3}$$

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

For each program, the procedure is invoked by the IDL command ".*run procedure_name*". The procedure will ask you various questions, such as where you want to put the observations, and the error characteristics of the observations and background field.

To retain the previous parameter settings just press ENTER; to alter a parameter enter the new value at the prompt. (Note that if you want to change an array, you will need to enter the correct number of new values.)

The procedure will then plot a graph showing the observations, background and analysis. You then have the option of writing a postscript file (for a hard copy of the plot).

Finally, you are given the option of repeating the calculation.

1 Analysis_2obs

This program calculates an analysis from two observations using the formula (1). The inputs, their definitions and the values that they are initially set to (you can change these later) are given in Table 1.

Variable	Description	Value
o1	First observation	1.0
x1	Position of first observation	4.0
o2	Second observation	-1.0
x2	Position of second observation	6.0
sigo	σ_o the observation error	1.0
Lo	Observation error correlation length scale	0.0
sigf	σ_f the background error	1.0
Lf	Background error correlation length scale	2.0

Table 1: Definitions of abbreviations and initial values

Observations are values between -2 and +2 and horizontal (x) positions are between 0 and 10

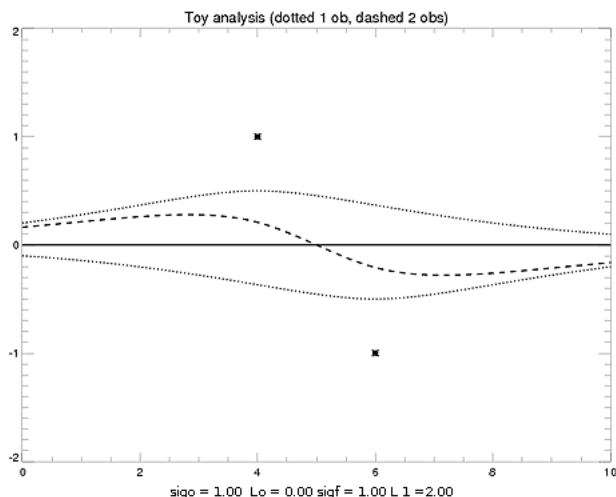The background in this procedure is set to be zero.

Figure 1: The output produced by Analysis_2obs with the data defined above in Table 1. The two dotted lines are the results if just one observation is used and the dashed line is the result from the two observations. The solid black line at zero is the background.

Figure 1 is the output of the program Analysis_2obs when the input data that is given to the procedure is that of the third column of Table 1. 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 and the solid black line at zero is the background. Try to run the program to reproduce the above graph (Fig. 1) by typing *.run Analysis_2obs* and then pressing ENTER at each option (this re-uses the previous value).

Now you can explore what happens if you change the inputs.

- What happens if you change the observation values and locations?

- What happens if you vary the background error correlation length Lf, (e.g. from 0. to 10.)? What happens if you vary the background error sigf, both increasing and decreasing?

- The effect of the observation error covariances is less obvious; what happens if the observation errors are highly correlated and the observations agree with one another?

- What happens if they disagree with one another?

Try to explain the results you find.

2 Analysis_mobs

This program calculates the analysis of two sets of observations, with a maximum of seven in each set. The initial set up of the observations:

SET A:

3 observations

Position	4.0	5.0	6.0
Value	1.0	1.0	1.0

Observation error (sigoa): 2.0

Observation error lengthscale (Loa): 0.0

SET B:

3 observations

Position	4.0	5.0	6.0
Value	-1.0	-1.0	-1.0

Observation error (sigob): 1.0

Observation error lengthscale (Lob): 2.0

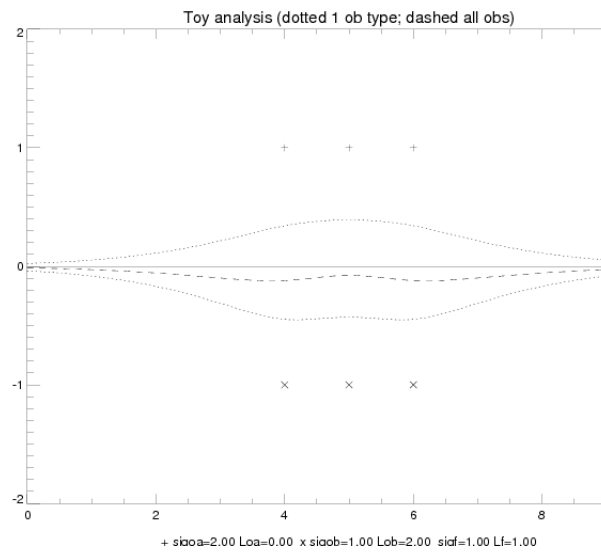The background is set to have an error (sigf) of 1.0 and a lengthscale (Lf) of 1.0

Figure 2: The output produced by Analysis_mobs with the initial data defined above. The dotted lines are the result if just one set of observations is used and the dashed line is the result if both sets of observations are used.

Again explore what happens if you change the inputs. You could look at:

- the effect of varying the number of observations, their error magnitudes, and their spatial density.

- 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.

3 Analysis_sim

This procedure simulates observations and the background field with specified error characteristics. You have several choices as to the "truth" field.

0 : Zero everywhere (initially the truth is set to be this)
1 : Psuedo step function
2 : Sum of several sinusoidal curves
3 : Statistically generated

The background is generated from the truth field, either by adding on random perturbations (according to the background error parameters), or by shifting the truth field along the x-axis. Initially the background has an error (sigf) of 2.0 and a lengthscale (Lf) of 2.0.

The observations are generated from the truth field, again using the specified observation error statistics. There can be upto 11 observations, initially there are 5 at positions: 3.0, 4.0 5.0, 6.0 and 7.0. The observation error (sigo) is 1.0 and the lengthscale (Lo) is 0.0
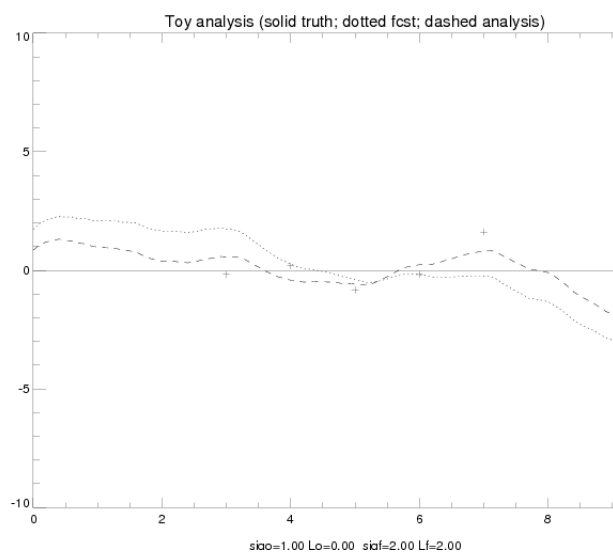

Figure 3: The output produced by Analysis sim with the initial data. The dotted line is the forecast and the dashed line is the analysis.

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.

- One 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?

4 Further exercises

The procedures as written are quite basic. Feel free to take copies and amend them to carry out your own explorations.

In analysis_sim, you can easily amend the code to explore what happens if the assumed observation and background errors are different from their true values.

4

Introduction to DA concepts

Debbie Clifford

National Centre for Earth Observation

University of Reading

What terms do I need to know?

- **Truth** – the thing we are trying to discover
- **Background** – the thing we know already
- **Observations** – the new information
- **Error covariance** – how much we believe either the background or the observations
- **Analysis** – our estimate at the end of the procedure
- **Dynamical model** – for tomorrow!

But what do they all *really* mean?

How would you estimate the temperature in this room?

➢ Make some observations

➢ Use some prior knowledge

➢ Assume some errors in both

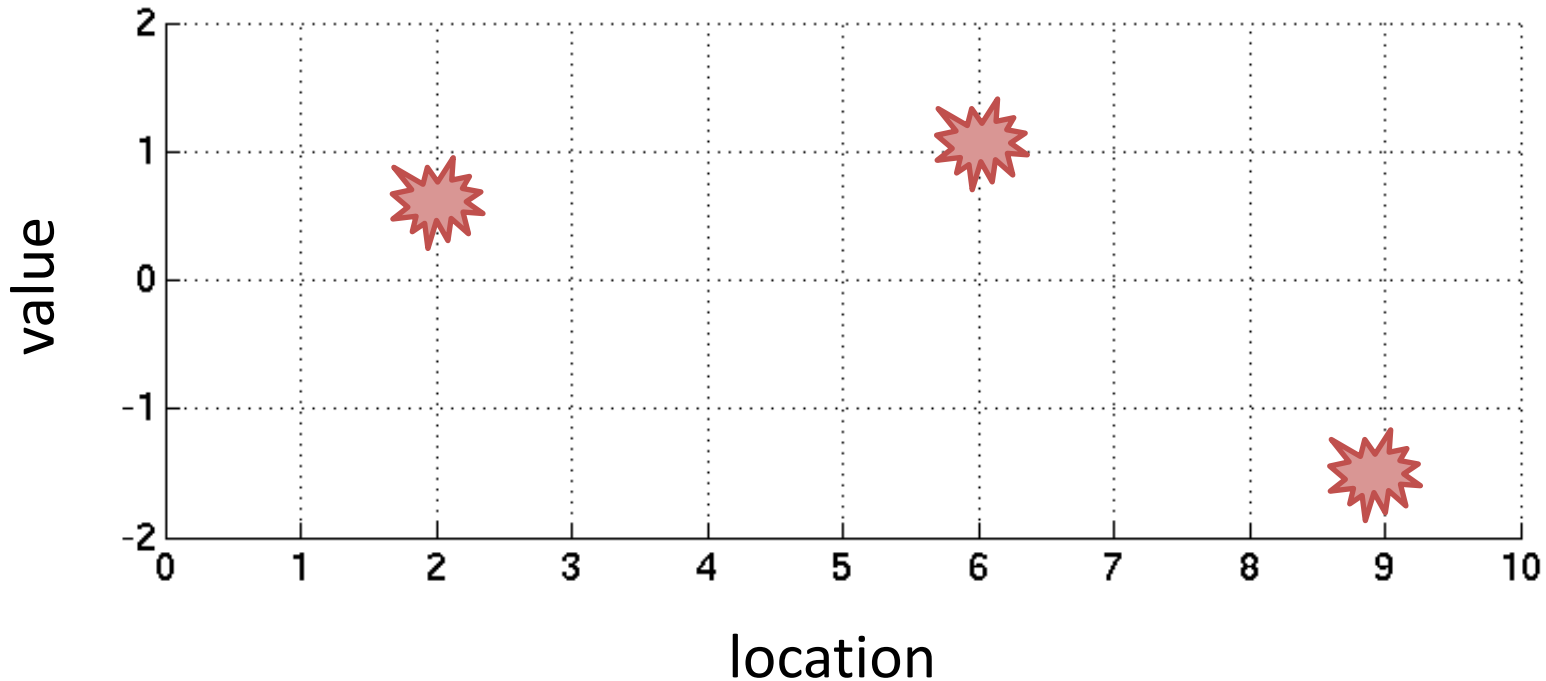➢ Combine these in some way to produce an answer – a 'best estimate'

Key questions…

- What happens if observations are close together?
- What if the observations are rubbish?
- How does the method cope with correlated errors?
- What if you believe the background, but it's actually wrong?
- What if I have some dense, inaccurate observations and some sparse but accurate ones?
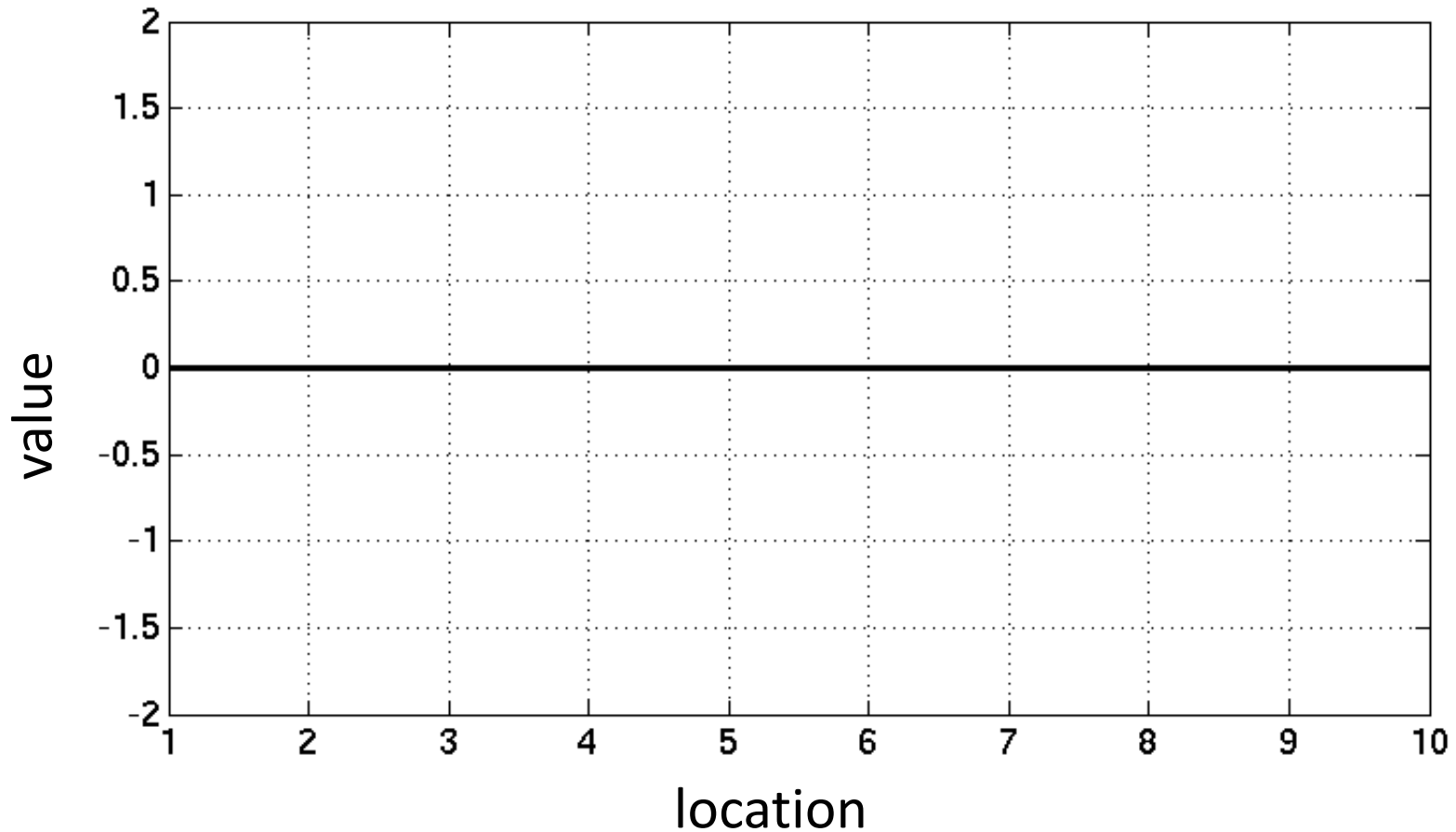- And so on…

➢ This is what we will explore in this practical

So in this practical we…

- **Make some observations**
- Use a background or forecast term
- Assume some errors in both
- Use an algorithm to combine these to produce an answer – a

So in this practical we…

- Make some observations
- Use a background or forecast term

So in this practical we…

- Make some observations
- Use a background or forecast term
- Assume some errors in both
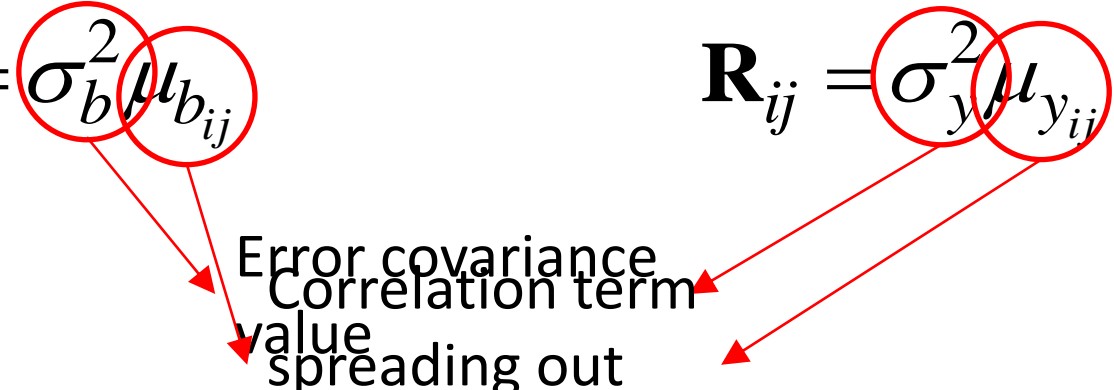- Use an algorithm to combine these to produce an answer – a 'best estimate'

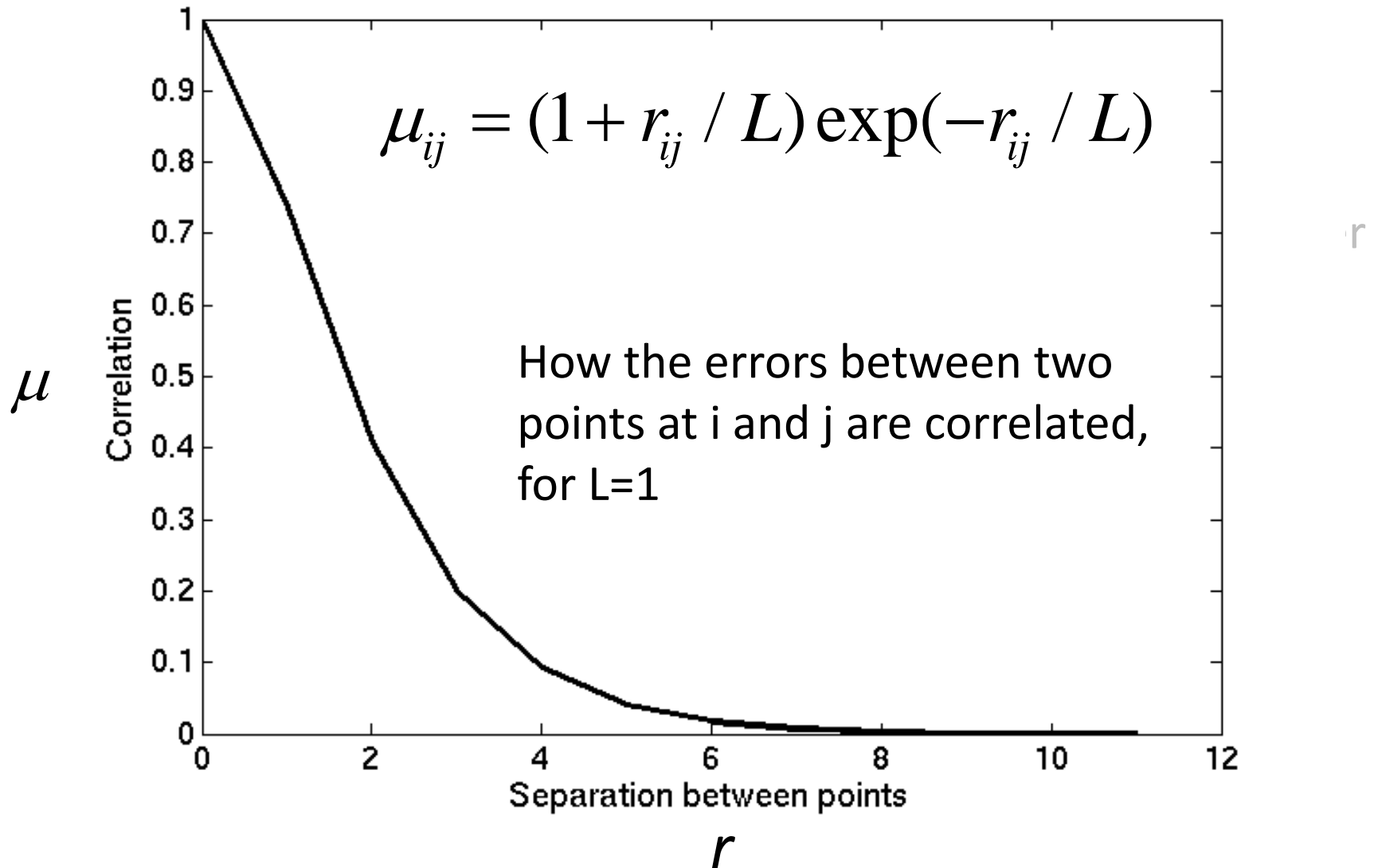Background:

$$\mathbf{B}_{ij} = \sigma_b^2 \mu_{b_{ij}}$$

Observations:

$$\mathbf{R}_{ij} = \sigma_y^2 \mu_{y_{ij}}$$

Error covariance value

Correlation term spreading out information

So in this practical we...

$$\mu_{ij} = (1 + r_{ij} / L)\exp(-r_{ij} / L)$$

How the errors between two points at i and j are correlated, for L=1

So in this practical we…

- Make some observations
- Use a background or forecast term
- Assume some errors in both

- Use an algorithm to combine these to produce an answer – a 'best estimate'

  ➢ This is the 'optimal interpolation' equation we will use:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{BH}(\mathbf{HBH}^T + \mathbf{R})^{-1}(\mathbf{y}_o - \mathbf{Hx}_b)$$

The analysis is
the background (first guess)
updated by the difference between the observed value and the
background at the observation point,
weighted by a term that depends on the assumed error covariances.

Routines to run today

- ANALYSIS_2OBS
  - Constructs 1-D analysis from 2 idealised obs with specified values and error characteristics

- ANALYSIS_MOBS
  - Constructs 1-D analysis from multiple idealised obs with specified values and error characteristics

- ANALYSIS_SIM - Toy Analysis System
  - Constructs 1-D analysis from simulated truth, forecast and observations with specified values and error characteristics