An introduction to data assimilation

Amos S. Lawless

Data Assimilation Research Centre
University of Reading

a.s.lawless@reading.ac.uk

@amoslawless

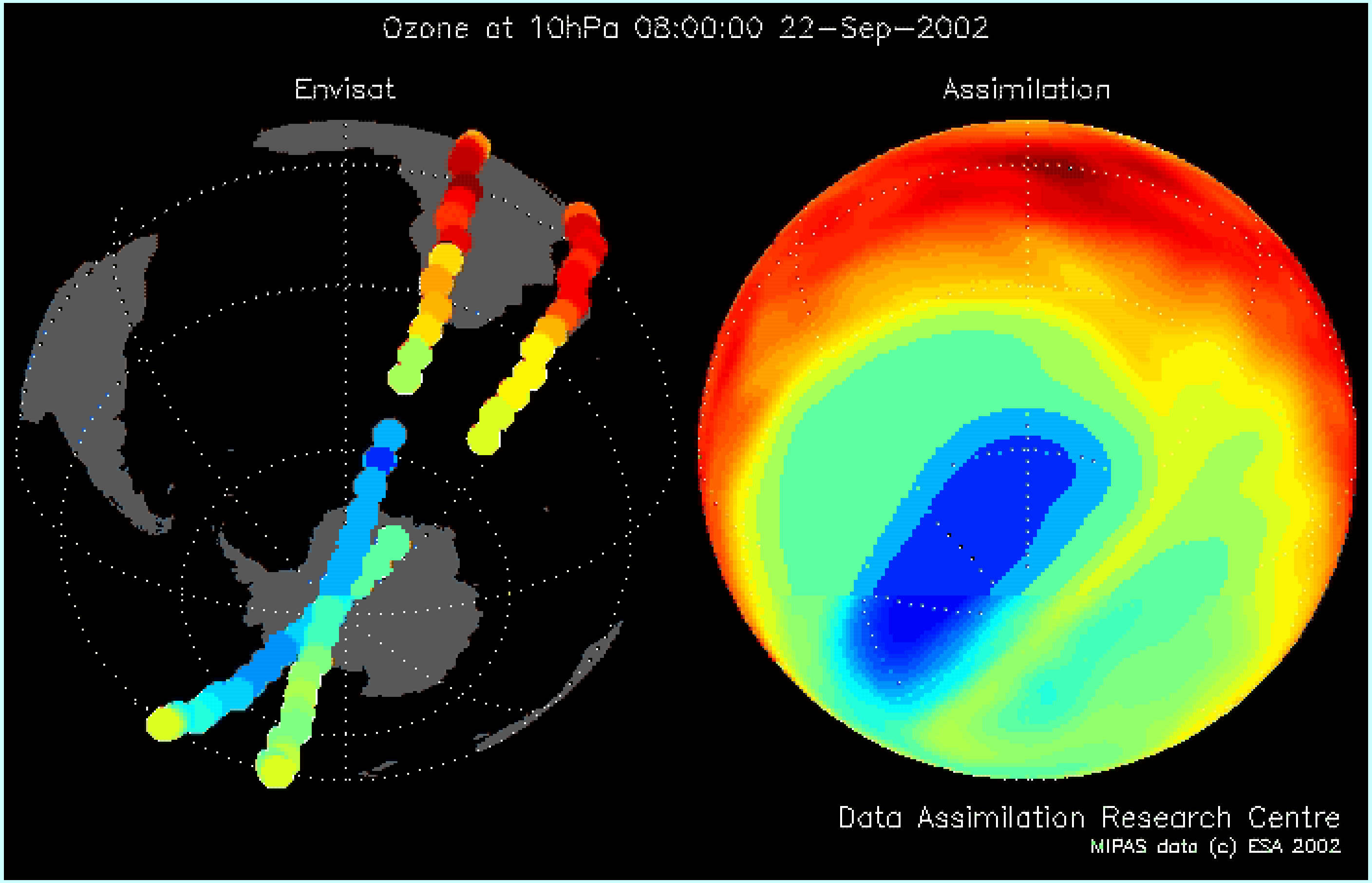
What is data assimilation?

Data assimilation is the process of estimating the state of a dynamical system by combining observational data with an *a priori* estimate of the state (often from a numerical model forecast).

We may also make use of other information such as

- The system dynamics
- Known physical properties
- Knowledge of uncertainties

Example – ozone hole

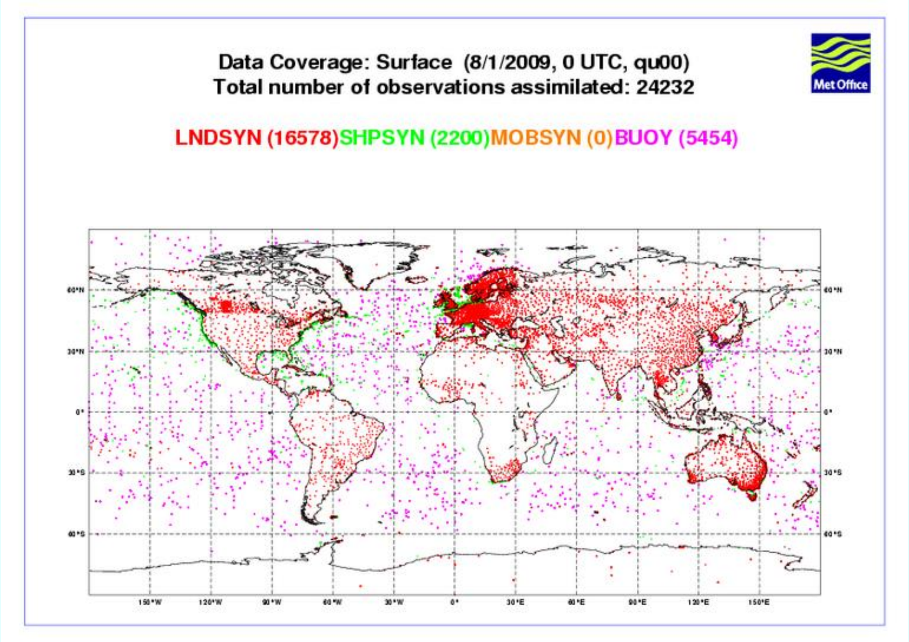


National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

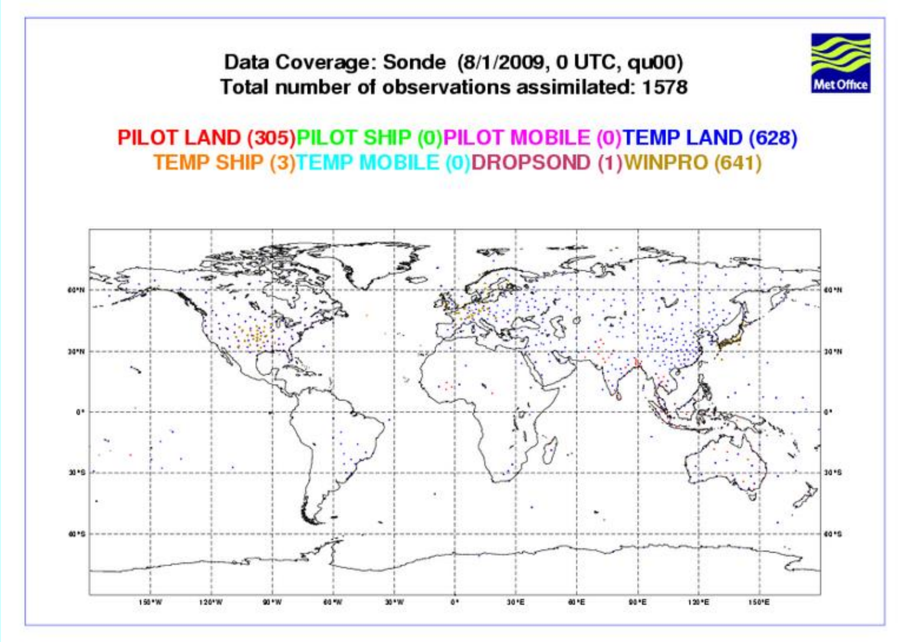
University of Reading

Why not just use the observations?

1. We may only observe part of the state



Surface



Radiosonde

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

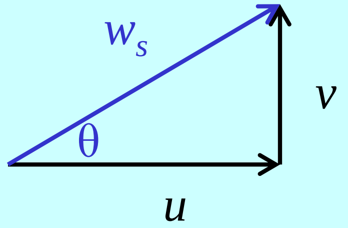
University of Reading

Why not just use the observations?

2. We may observe a nonlinear function of the state, e.g. satellite radiances.

Example

Let the state vector consists of the E-W and N-S components of the wind, *u* and *v*.



Suppose we observe the wind speed w_s.

Then we have $\mathbf{x} = \begin{pmatrix} u \\ v \end{pmatrix}$, $\mathbf{y} = w_s$ and $\mathbf{y} = H(\mathbf{x})$

with

$$H(\mathbf{x}) = \sqrt{u^2 + v^2}$$

H is known as the observation operator.

Why not just use the observations?

3. We need to allow for uncertainties in the observations (and in the *a priori* estimate).

A scalar example

Suppose we have a background estimate of the temperature in this room T_b and a measurement of the temperature T_o.

We assume that these estimates are unbiased and uncorrelated.

What is our best estimate of the true temperature?

We consider our best estimate (analysis) to be a linear combination of the background and measurement

$$T_a = \alpha_b T_b + \alpha_o T_o$$

Then the question is how should we choose α_b and α_o?

We need to impose 2 conditions.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

1. We want the analysis to be unbiased.

Let

$$T_a = T_t + \epsilon_a$$
$$T_b = T_t + \epsilon_b$$
$$T_o = T_t + \epsilon_o$$

Then

$$
\begin{aligned}
< \epsilon_a > &= < T_a - T_t > \\
&= < \alpha_b T_b + \alpha_o T_o - T_t > \\
&= < \alpha_b (T_b - T_t) + \alpha_o (T_o - T_t) + (\alpha_b + \alpha_o - 1) T_t > \\
&= \alpha_b < \epsilon_b > + \alpha_o < \epsilon_o > + (\alpha_b + \alpha_o - 1) < T_t >
\end{aligned}
$$

Hence to ensure that $< \epsilon_a > = 0$ for all values of T_t we require that

$$\alpha_b + \alpha_o = 1$$

so

$$T_a = \alpha_b T_b + (1 - \alpha_b) T_o$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

2. We want the uncertainty in our analysis to be as small as possible, i.e. we want to minimize its variance

Let

$$< \epsilon_b^2 > = \sigma_b^2$$
$$< \epsilon_o^2 > = \sigma_o^2$$
$$< \epsilon_a^2 > = \sigma_a^2$$

Then

$$\begin{aligned}\sigma_a^2 &= < (T_a - T_t)^2 > \\ &= < (\alpha_b T_b + (1-\alpha_b)T_o - T_t)^2 > \\ &= < (\alpha_b(T_b - T_t) + (1-\alpha_b)(T_0 - T_t))^2 > \\ &= < \ (\alpha_b\,\epsilon_b + (1-\alpha_b)\epsilon_o)^2 > \\ &= \alpha_b^2\sigma_b^2 + (1-\alpha_b)^2\sigma_o^2\end{aligned}$$

using $< \epsilon_b \epsilon_o > = 0$

Then setting $\dfrac{d\sigma_a^2}{d\alpha_b} = 0$ we find

$$\alpha_b = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Hence we have

$$T_a = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2} T_b + \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2} T_o$$

This is known as the Best Linear Unbiased Estimate (BLUE).

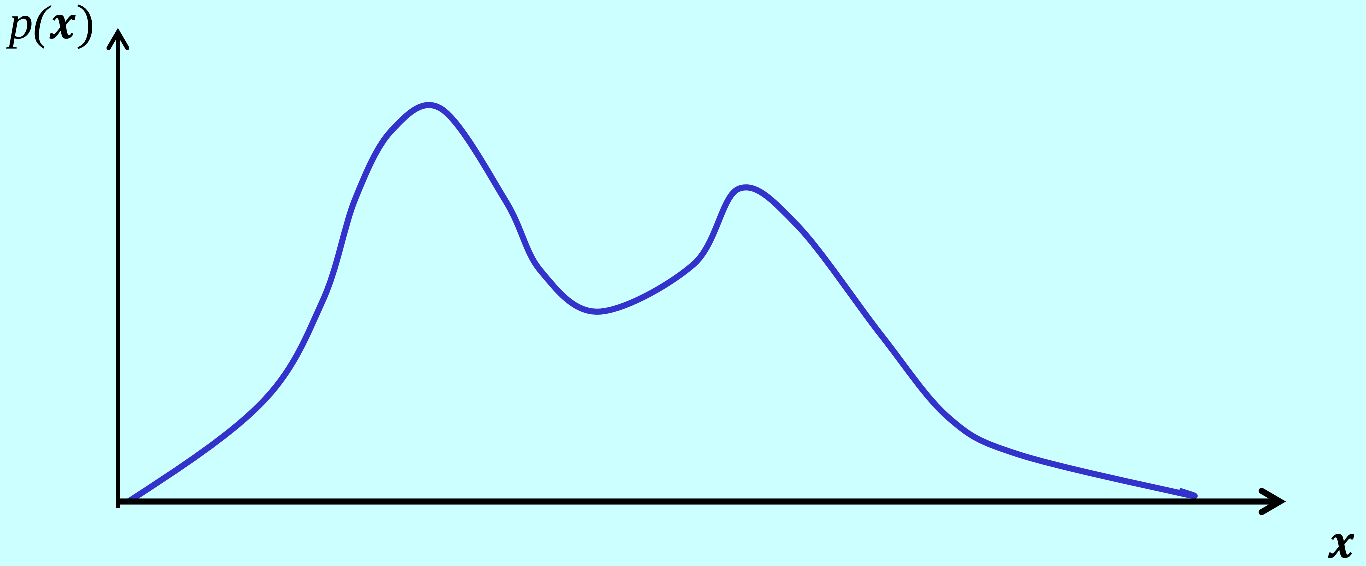
We find that

$$\sigma_a^2 = \frac{\sigma_b^2 \sigma_o^2}{\sigma_b^2 + \sigma_o^2} < \min\{\sigma_b^2, \sigma_o^2\}$$

How can we generalise this to a vector state and a vector of observations?

More general problem

In order to generalise the problem we need to use probability distribution functions (pdf's) to represent the uncertainty.



Bayes theorem

We assume that we have

- A prior distribution of the state **x** given by $p(\mathbf{x})$
- A vector of observations **y** with conditional probability $p(\mathbf{y}|\mathbf{x})$

Then Bayes theorem states

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Prior

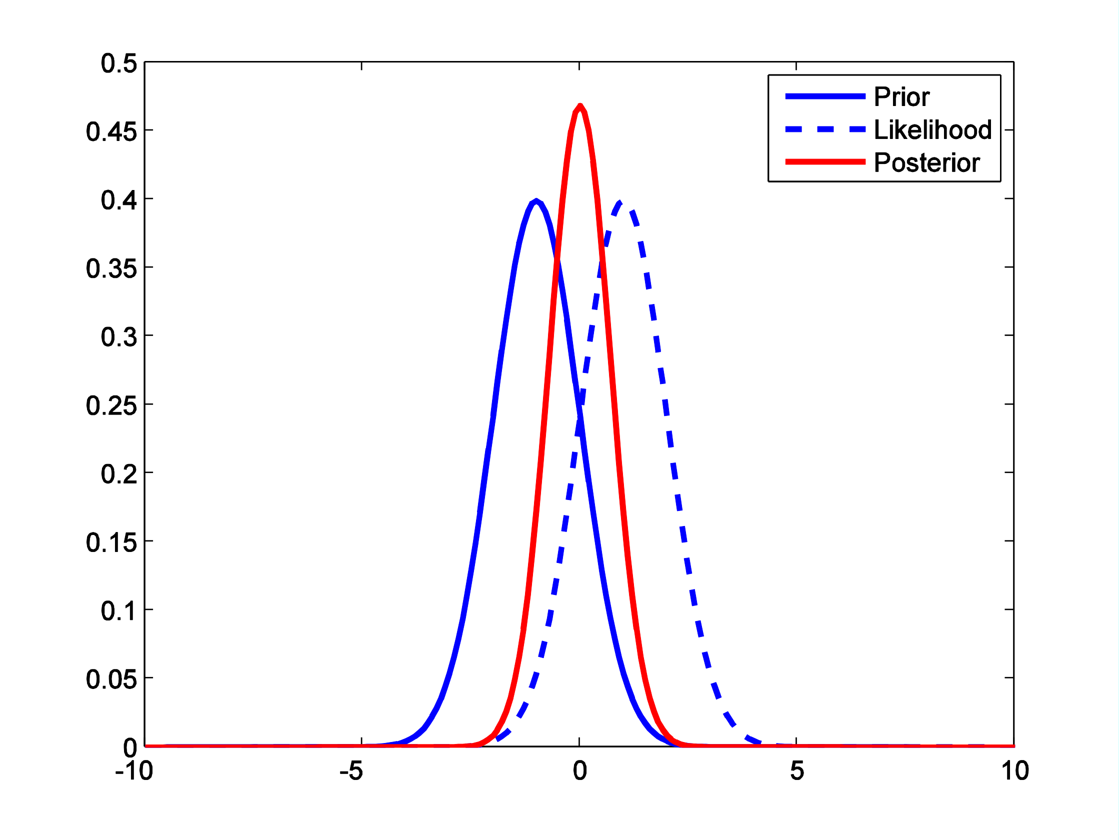
Likelihood

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

Posterior

Normalising constant

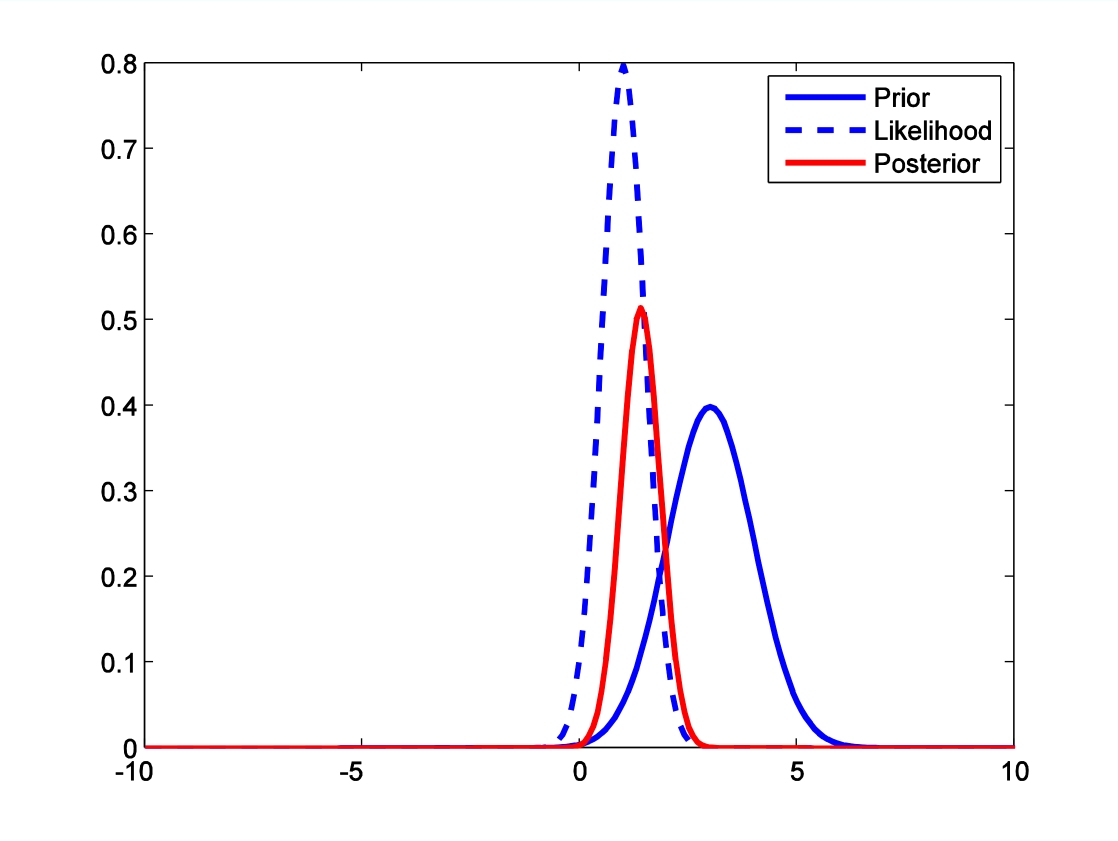
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

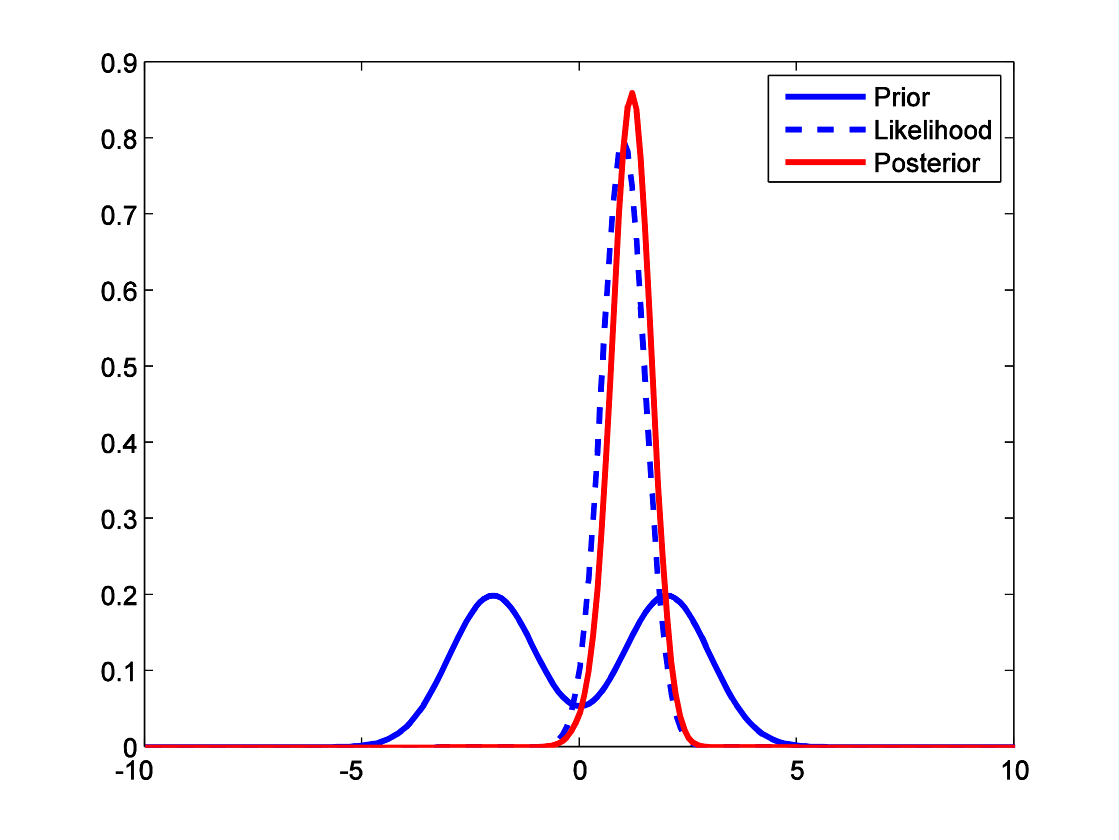
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$



But … In practice the pdf's are very high dimensional (e.g. 10^9 in NWP).

This means

- We cannot calculate the full pdf.
- We need to either calculate an estimator based on the pdf or generate samples from the pdf.

Gaussian assumption

If we assume that the errors are Gaussian then the pdf is defined solely by the mean and covariance.

Prior

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_b)\}$$

Likelihood

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{R}|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))\}$$

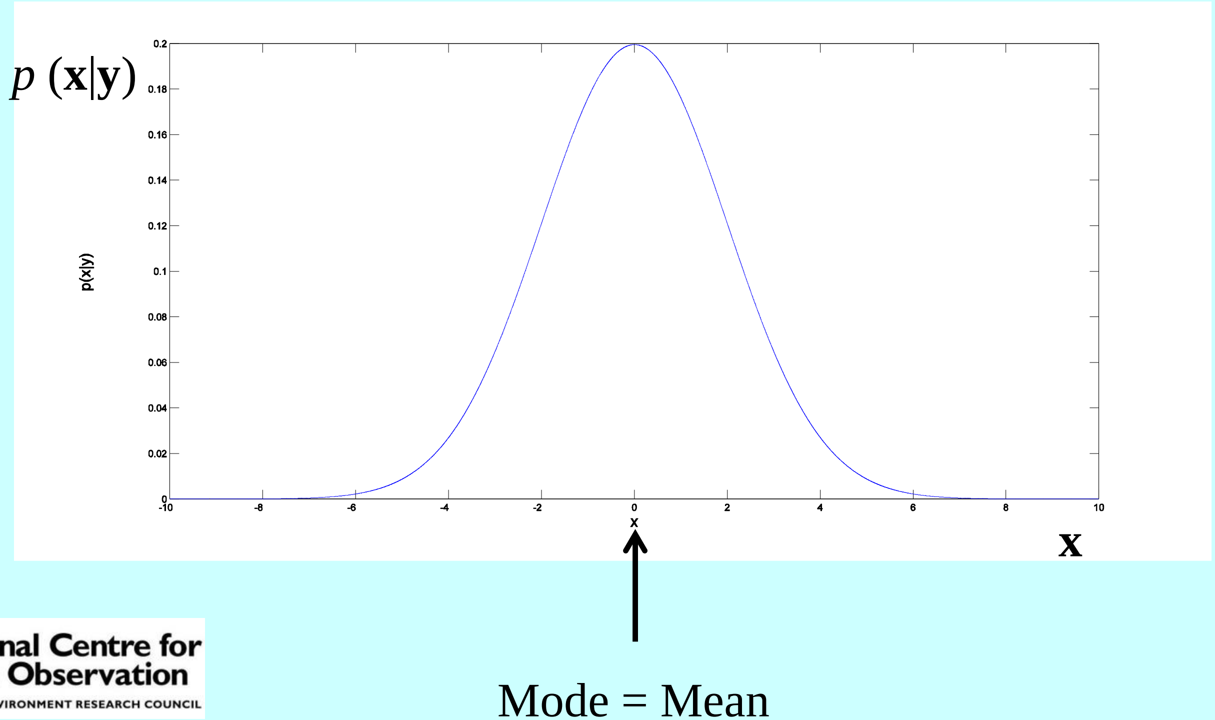
Posterior

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\{-\frac{1}{2}\{(\mathbf{x}-\mathbf{x}_b)^T\mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))\}\}$$

Maximum a posterior probability (MAP)

Find the state that is equal to the mode of the posterior pdf.

For a Gaussian case this is also equal to the mean.



Recall for the Gaussian case

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\{ -\frac{1}{2}\{(\mathbf{x}-\mathbf{x}_b)^T\mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))\}\}$$

So the maximum probability occurs when **x** minimises

$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^T\mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))$$

In the case of *H* linear we have

$$\mathbf{x} = \mathbf{x}_b + \mathbf{P}^T\mathbf{H}^T(\mathbf{HPH}^T + \mathbf{R})^{-1}(\mathbf{y}-H(\mathbf{x}_b))$$

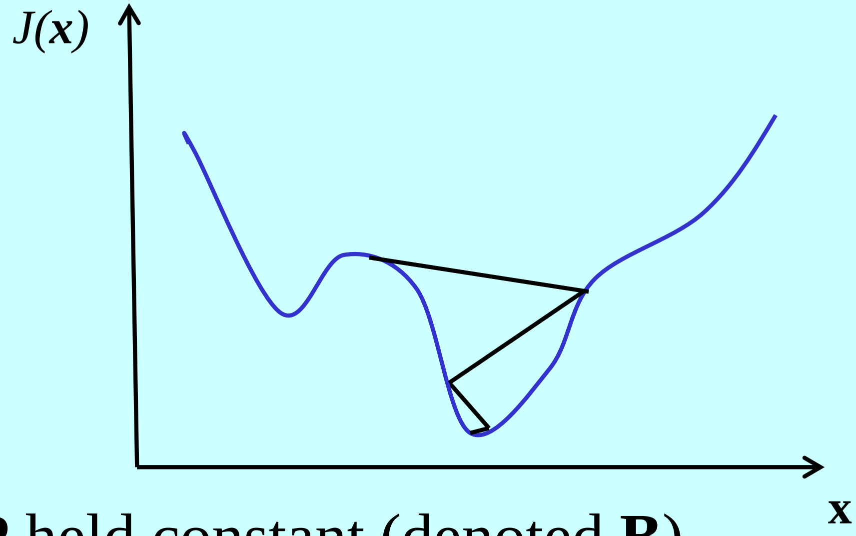
Note size of matrices!

How can we solve this in practice?

1. Variational methods (Ross Bannister, today)

Use an iterative optimization method to minimize

$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^T\mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))$$



Need gradient

$$\nabla J(\mathbf{x})$$

Usually **P** held constant (denoted **B**).

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

2. Kalman filter

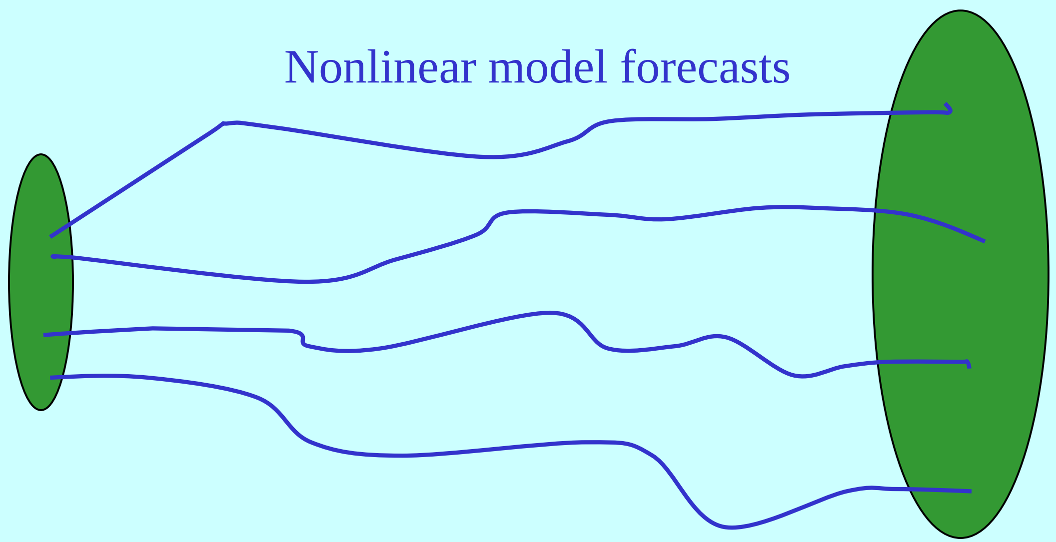
Solves directly

$$\mathbf{x} = \mathbf{x}_b + \mathbf{P}^T\mathbf{H}^T(\mathbf{HPH}^T + \mathbf{R})^{-1}(\mathbf{y} - H(\mathbf{x}_b))$$

- Only exact for linear case.
- Include update of covariance matrix **P** as system evolves.
- Can be extended to nonlinear case by linearization.

3. Ensemble Kalman filter (Alison Fowler & Sanita Vetra-Carvalho, Wed)

Similar to standard Kalman filter, but uses ensemble of nonlinear model runs to update covariance **P** at each assimilation time.



Uncertainty at analysis time

Uncertainty at forecast time with covariance **P** (Gaussian)

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

4. Hybrid methods (Javier Amezcua, Thurs)

A combination of variational and ensemble methods.

The covariances are generated from an ensemble, similar to the ensemble Kalman filter, but covariances are then used within a variational assimilation minimization problem.

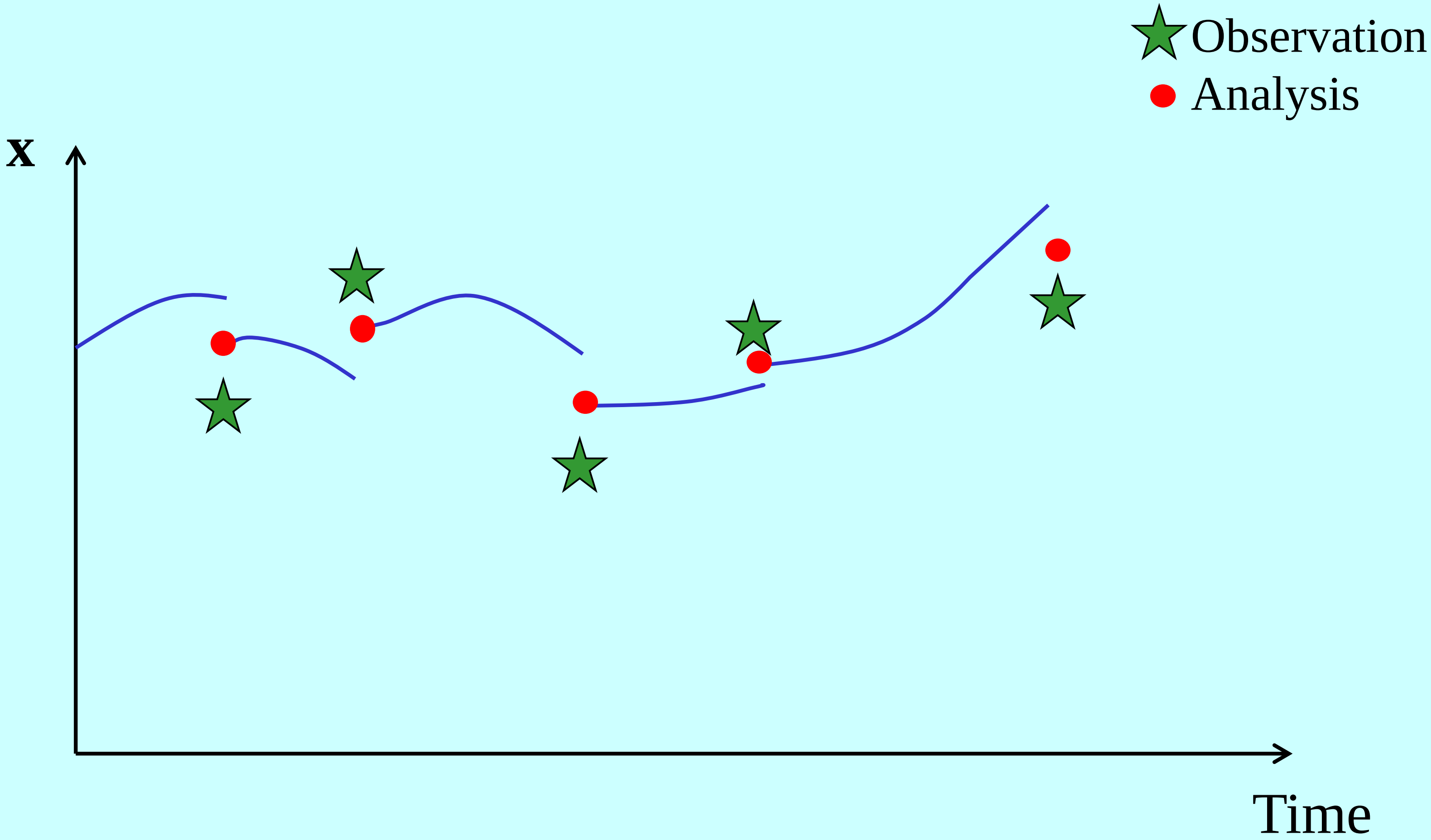
5. Particle filters (Javier Amezcua & Polly Smith, Fri)

Use a weighted sample of states to sample the true posterior pdf $p(\mathbf{x}|\mathbf{y})$.

As in Ensemble Kalman filter we use an ensemble of forecasts from the nonlinear model, but without making the Gaussian assumption.

Time sequence of observations

Filter – Treat observations sequentially in time

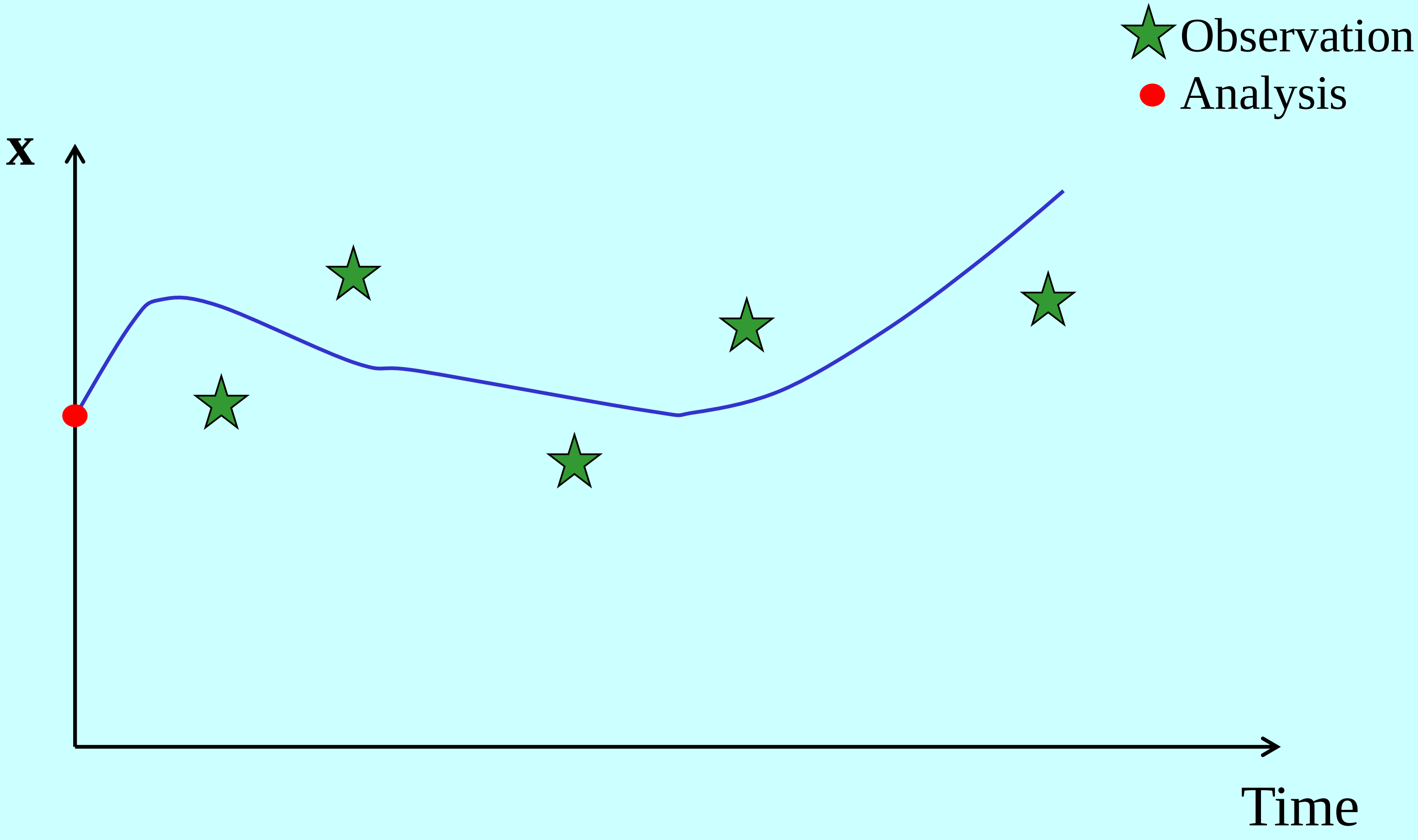


National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Time sequence of observations

Smoother – Treat all observations together



Summary

- Data assimilation provides the best way of using data with numerical models, taking into account what we know (uncertainty, physics, …).
- Bayes' theorem is a natural way of expressing the problem in theory.
- Dealing with the problem in practice is more challenging … This is the story of the rest of the week!

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading