

# An introduction to data assimilation

Amos S. Lawless

*a.s.lawless@reading.ac.uk*

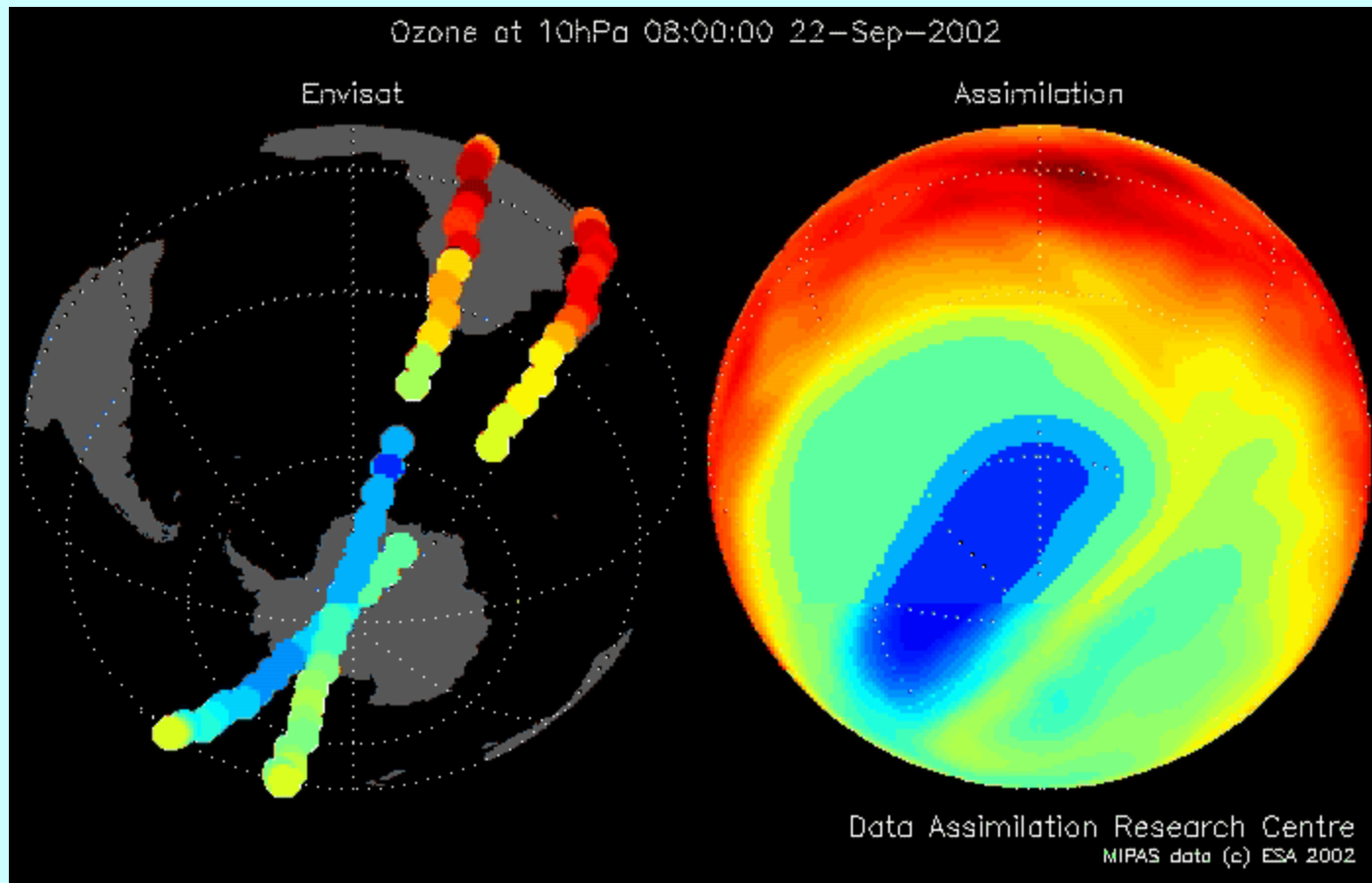
## What is data assimilation?

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

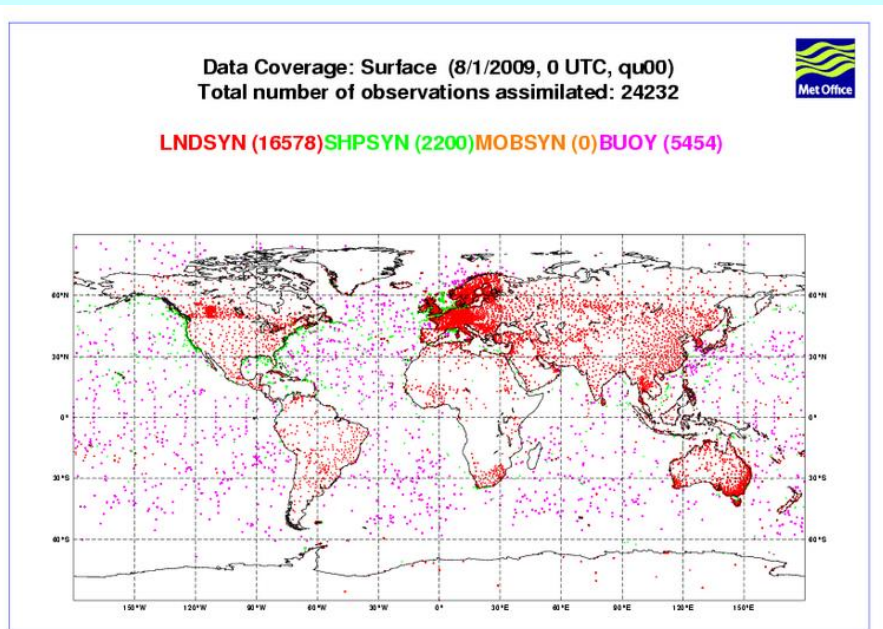
- The system dynamics
- Known physical properties
- Knowledge of uncertainties

# Example – ozone hole

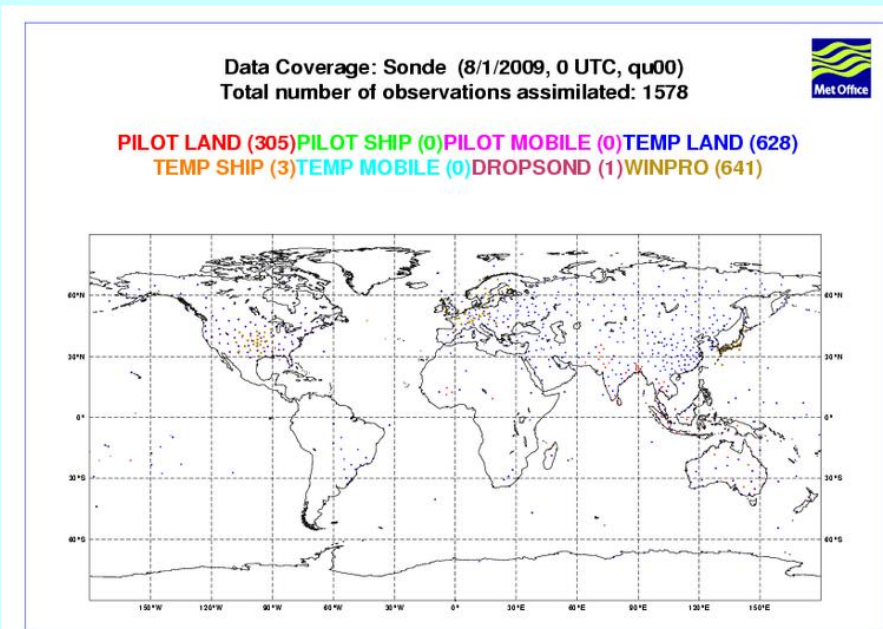


# Why not just use the observations?

## 1. We may only observe part of the state



Surface



Radiosonde

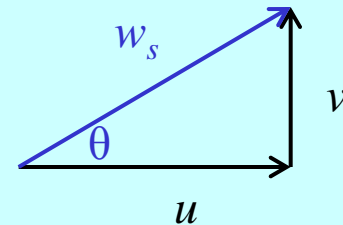
## Why not just use the observations?

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  $w_s$  and direction  $\theta$ .



Then we have

$$\text{State vector } \mathbf{x} = \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{Observation vector } \mathbf{y} = \begin{pmatrix} w_s \\ \theta \end{pmatrix}$$

These are related by the equation  $\mathbf{y} = H(\mathbf{x})$  where

$$H(\mathbf{x}) = \begin{pmatrix} \sqrt{u^2 + v^2} \\ \tan^{-1}\left(\frac{u}{v}\right) \end{pmatrix}$$

$H$  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  $\alpha_b$  and  $\alpha_o$ ?

We need to impose 2 conditions.



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

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

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

$$\alpha_b + \alpha_o = 1$$

so

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

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

Let

$$\langle \epsilon_b^2 \rangle = \sigma_b^2$$

$$\langle \epsilon_o^2 \rangle = \sigma_o^2$$

$$\langle \epsilon_a^2 \rangle = \sigma_a^2$$

Then

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

using  $\langle \epsilon_b \epsilon_o \rangle = 0$

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

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

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**).

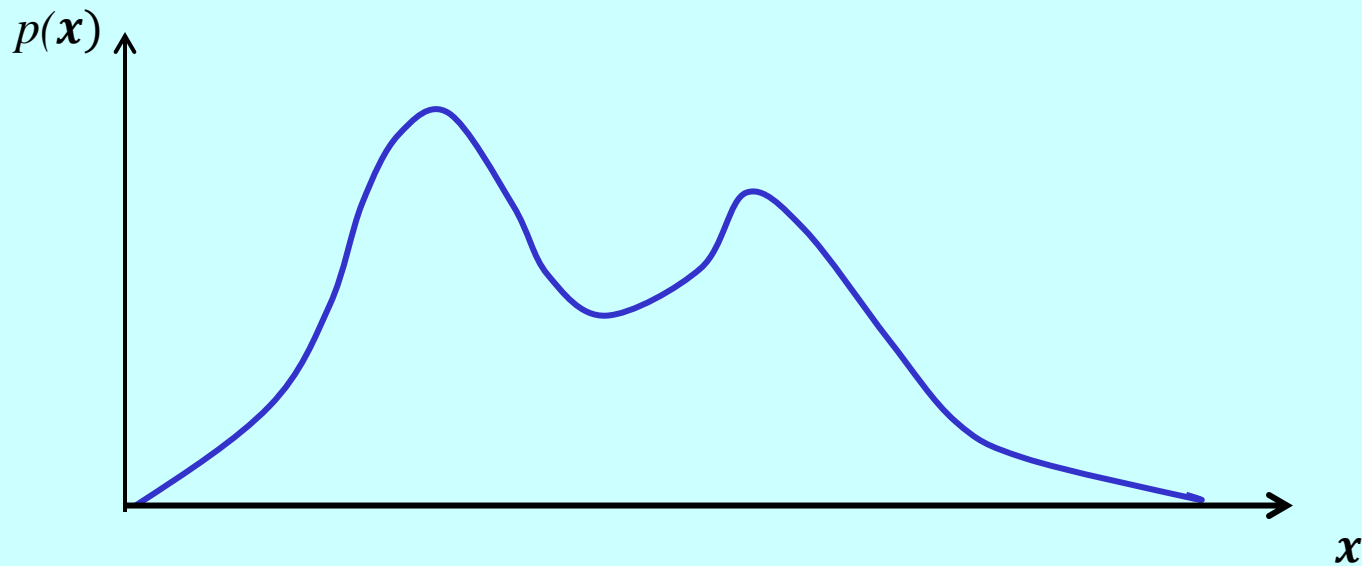
We find that

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

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.



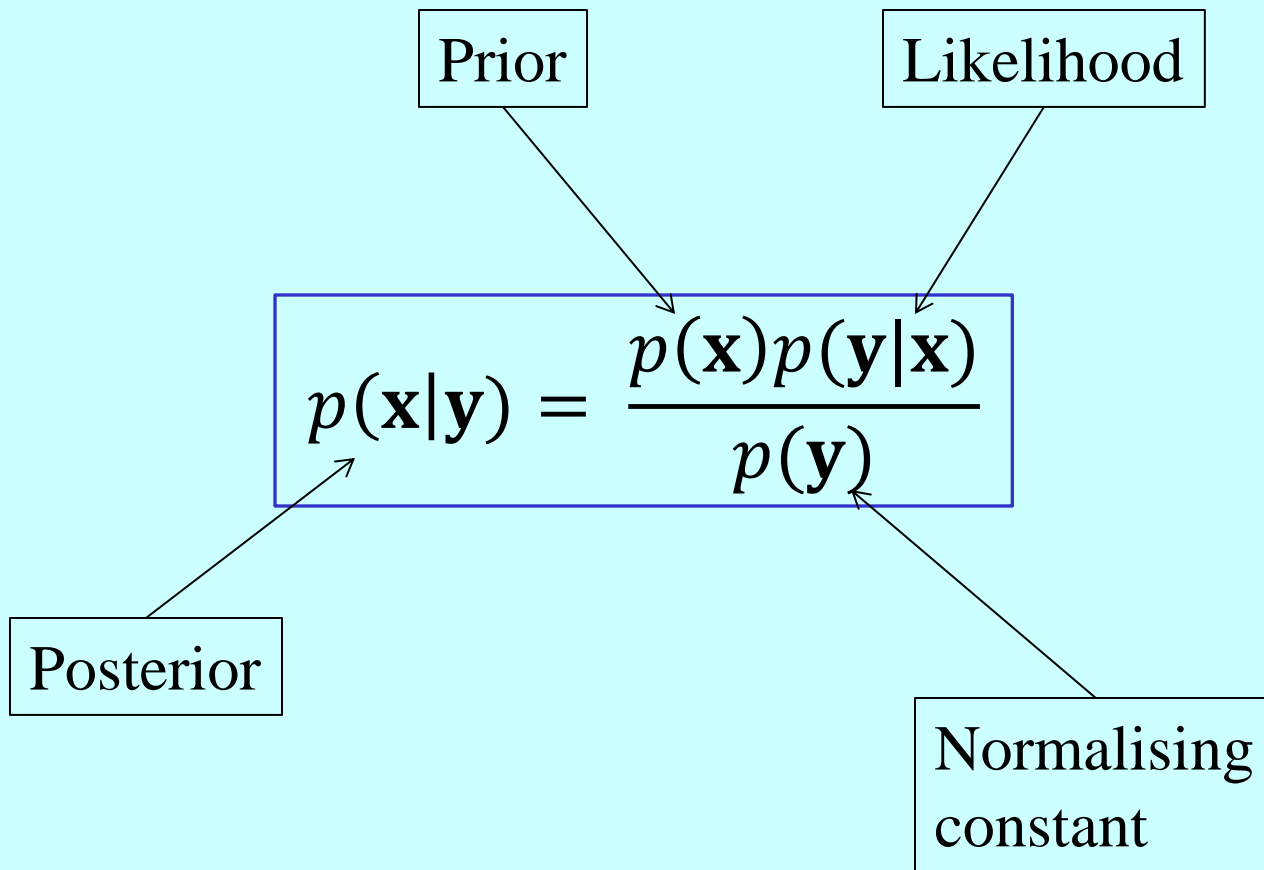
# Bayes theorem

We assume that we have

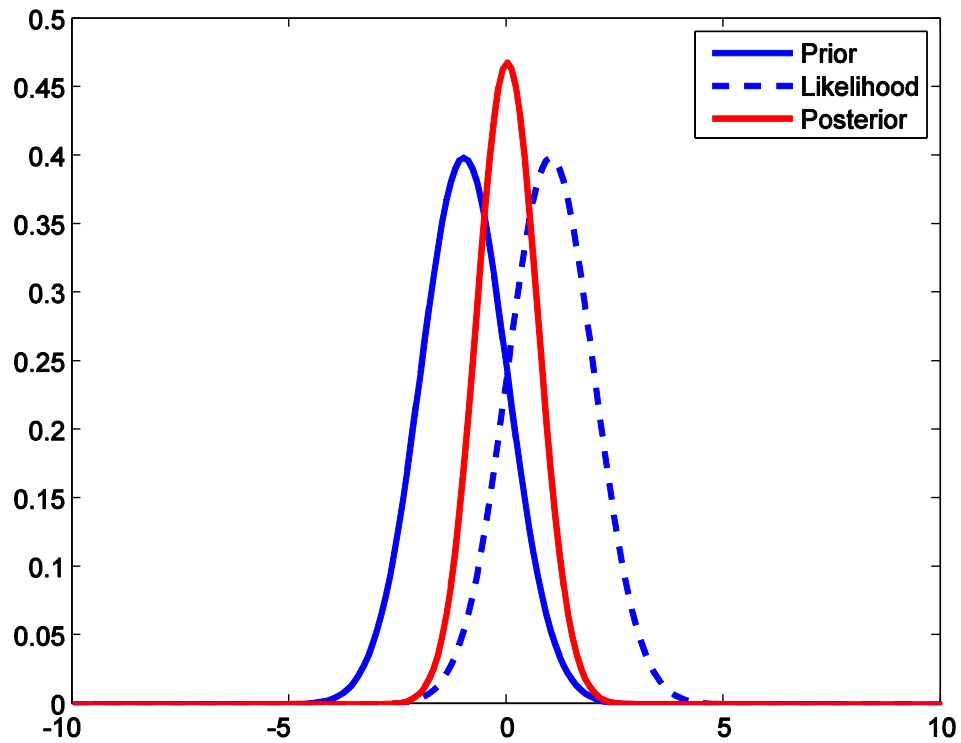
- A prior distribution of the state  $\mathbf{x}$  given by  $p(\mathbf{x})$
- A vector of observations  $\mathbf{y}$  with conditional probability  $p(\mathbf{y}|\mathbf{x})$

Then Bayes theorem states

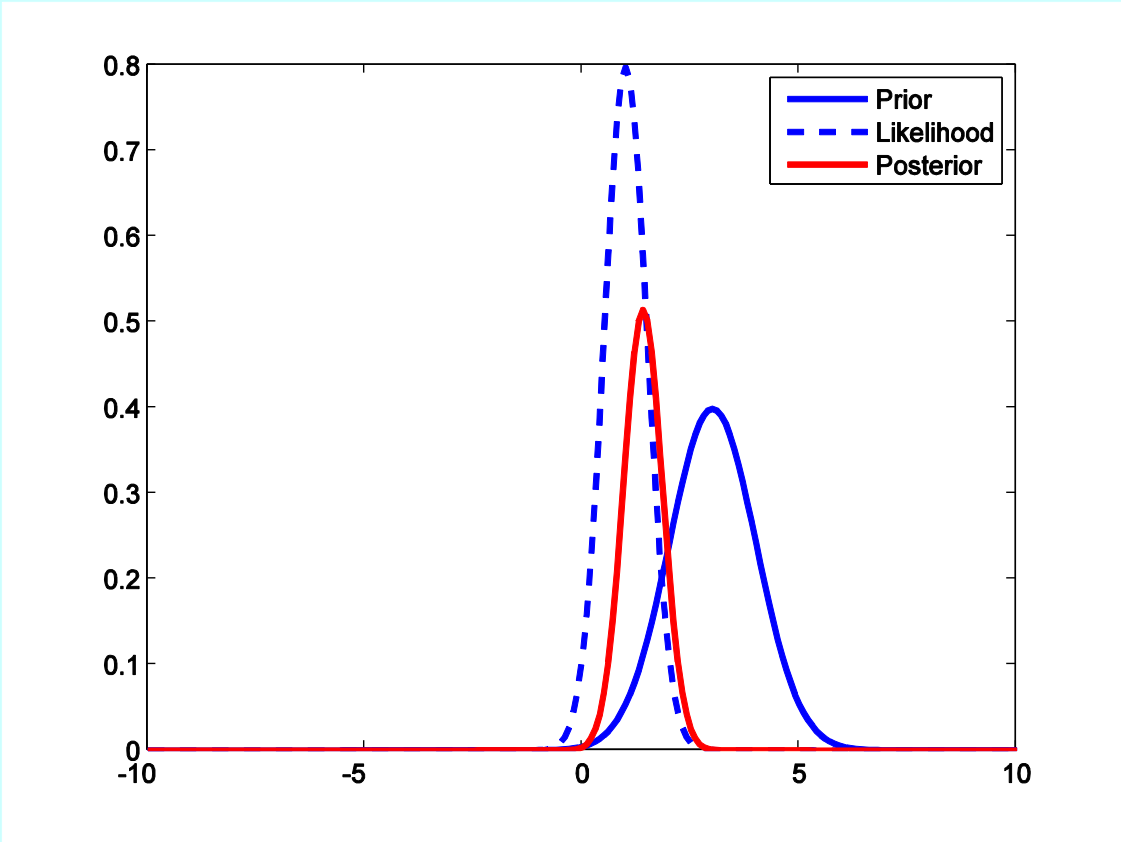
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$



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

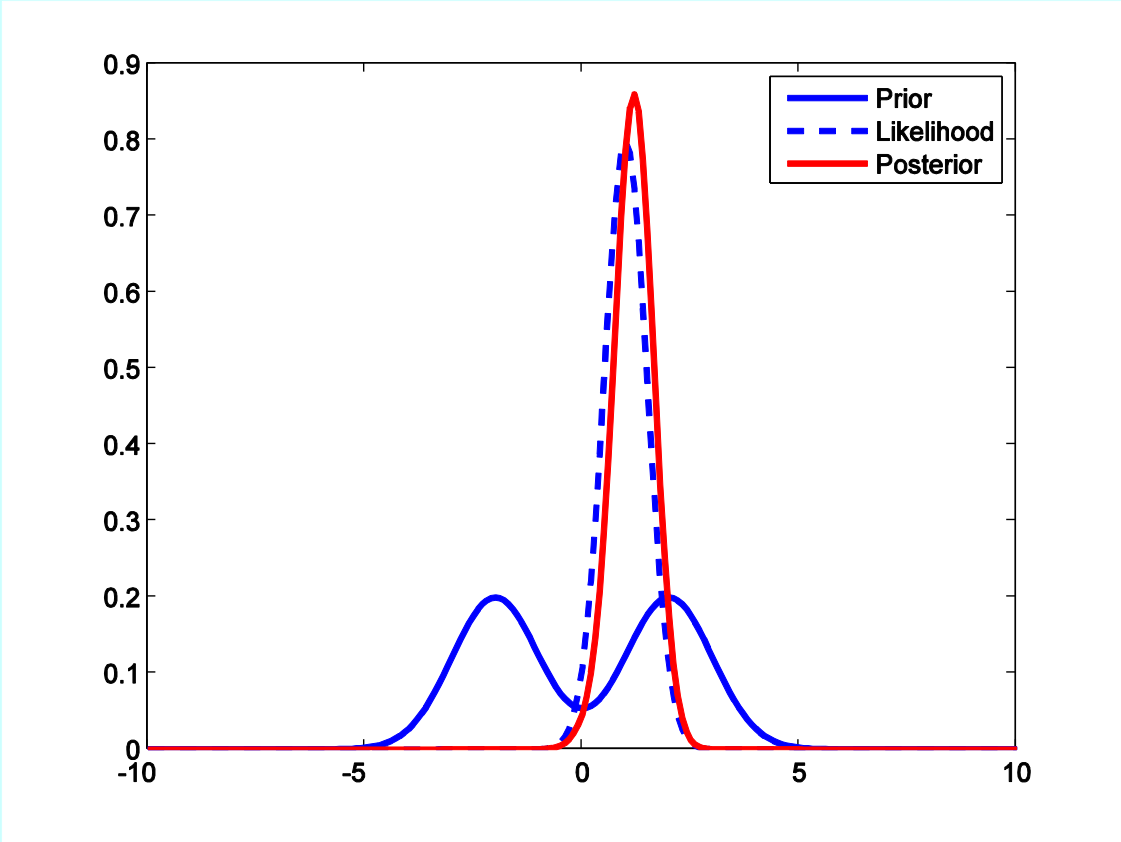


$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$





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



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

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

Likelihood

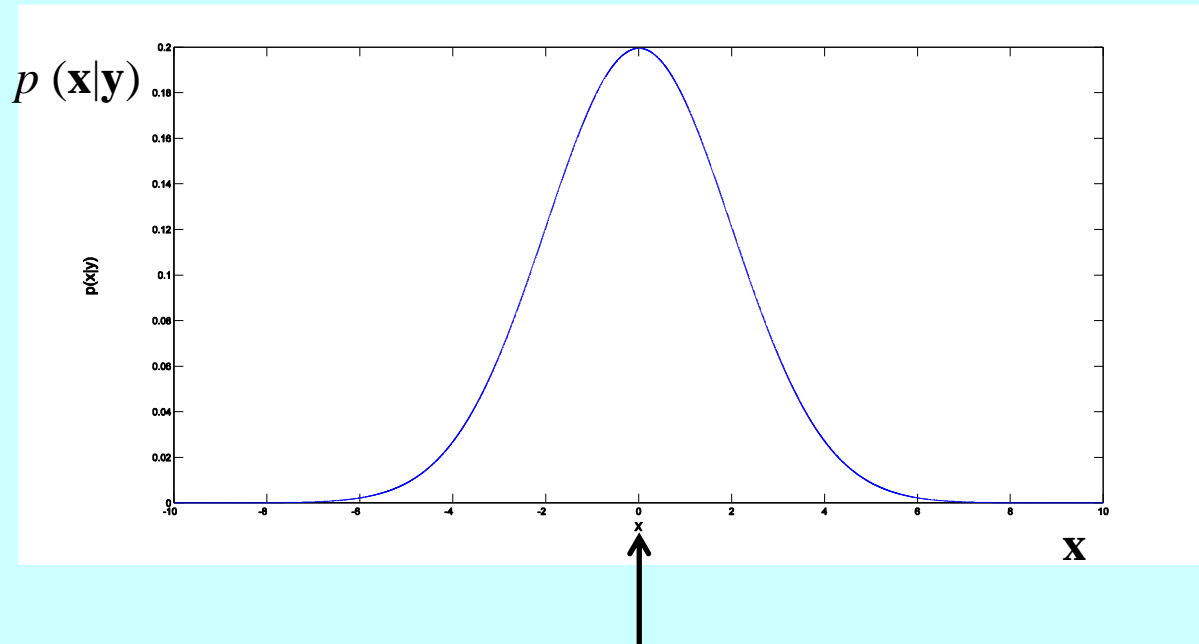
$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\mathbf{R}|^{p/2}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x})) \right\}$$

Posterior

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

## 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.



Mode = Mean

Recall for the Gaussian case

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

So the maximum probability occurs when  $\mathbf{x}$  minimises

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{P}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))$$

In the case of  $H$  linear we have

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

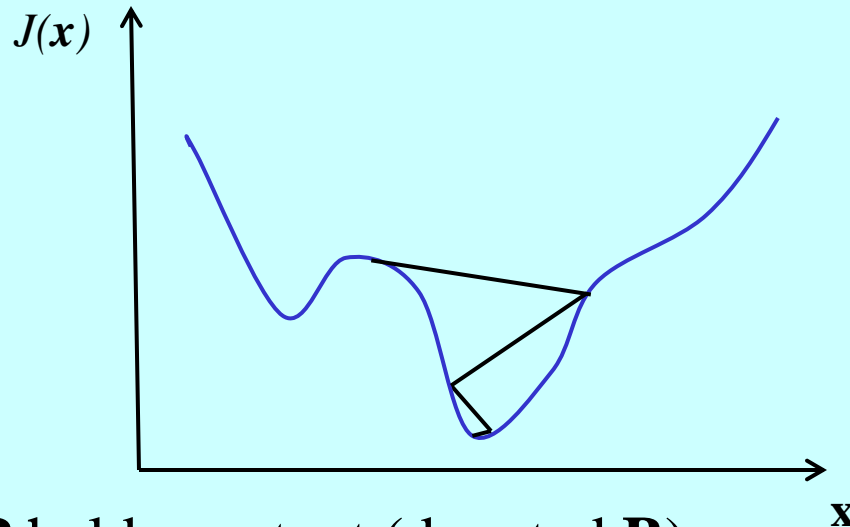
Note size of matrices!

# How can we solve this in practice?

## 1. Variational methods

Use an iterative optimization method to minimize

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{P}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))$$



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

Usually  $\mathbf{P}$  held constant (denoted  $\mathbf{B}$ ).

## 2. Kalman filter

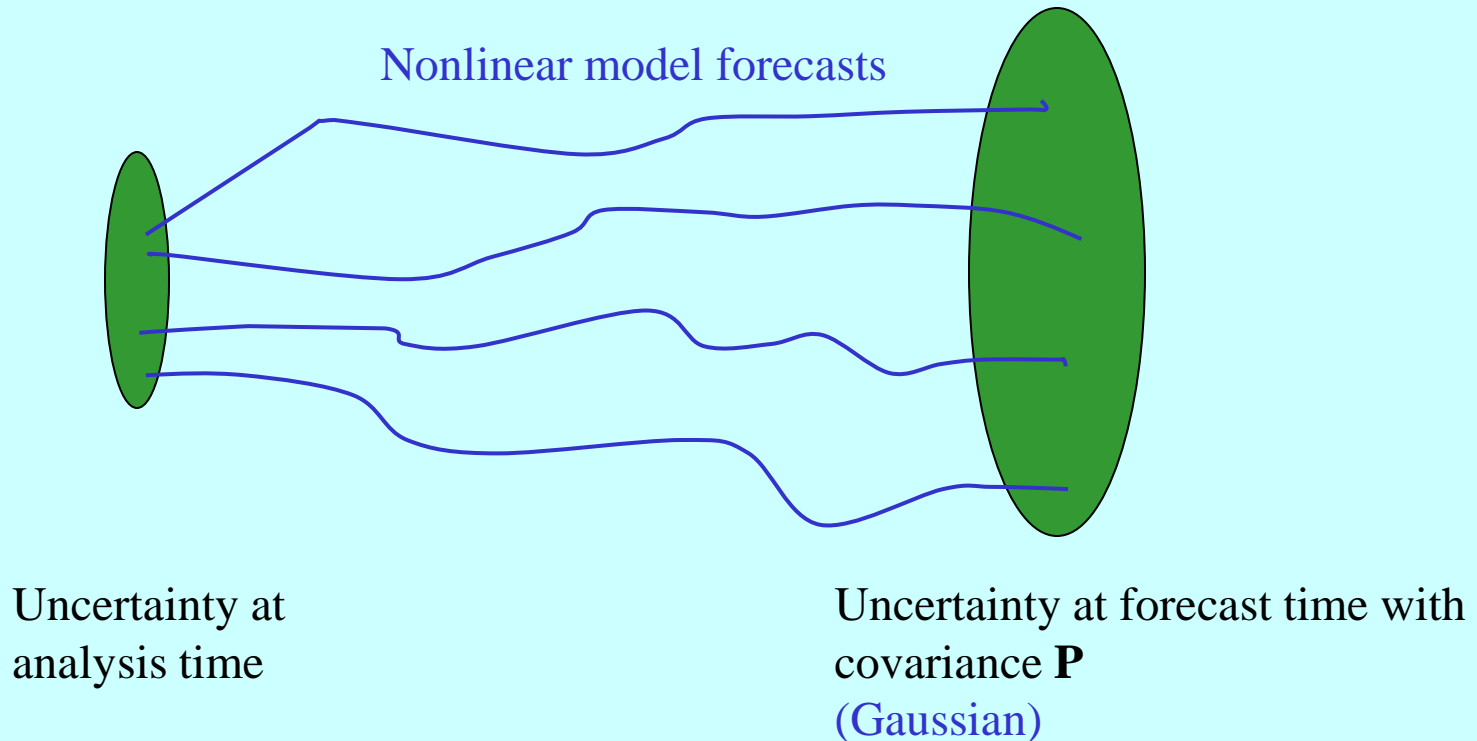
Solves directly

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

- Only exact for linear case.
- Include update of covariance matrix  $\mathbf{P}$  as system evolves.
- Can be extended to nonlinear case by linearization.

### 3. Ensemble Kalman filter

Similar to standard Kalman filter, but uses ensemble of nonlinear model runs to update covariance  $\mathbf{P}$  at each assimilation time.





## 4. Particle filters

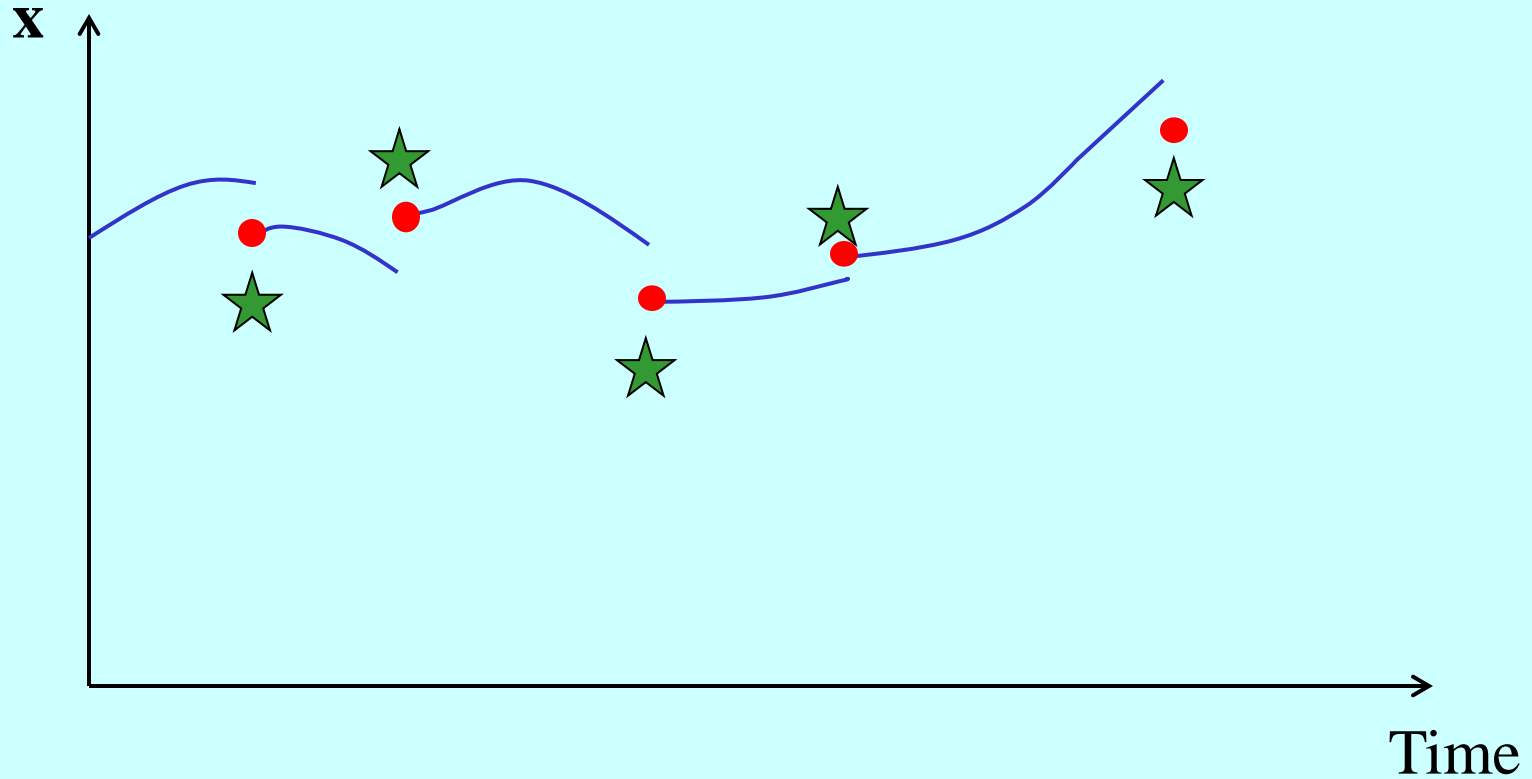
Use a weighted sample of states to sample the true posterior pdf  $p(\mathbf{x}|\mathbf{y})$ .

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

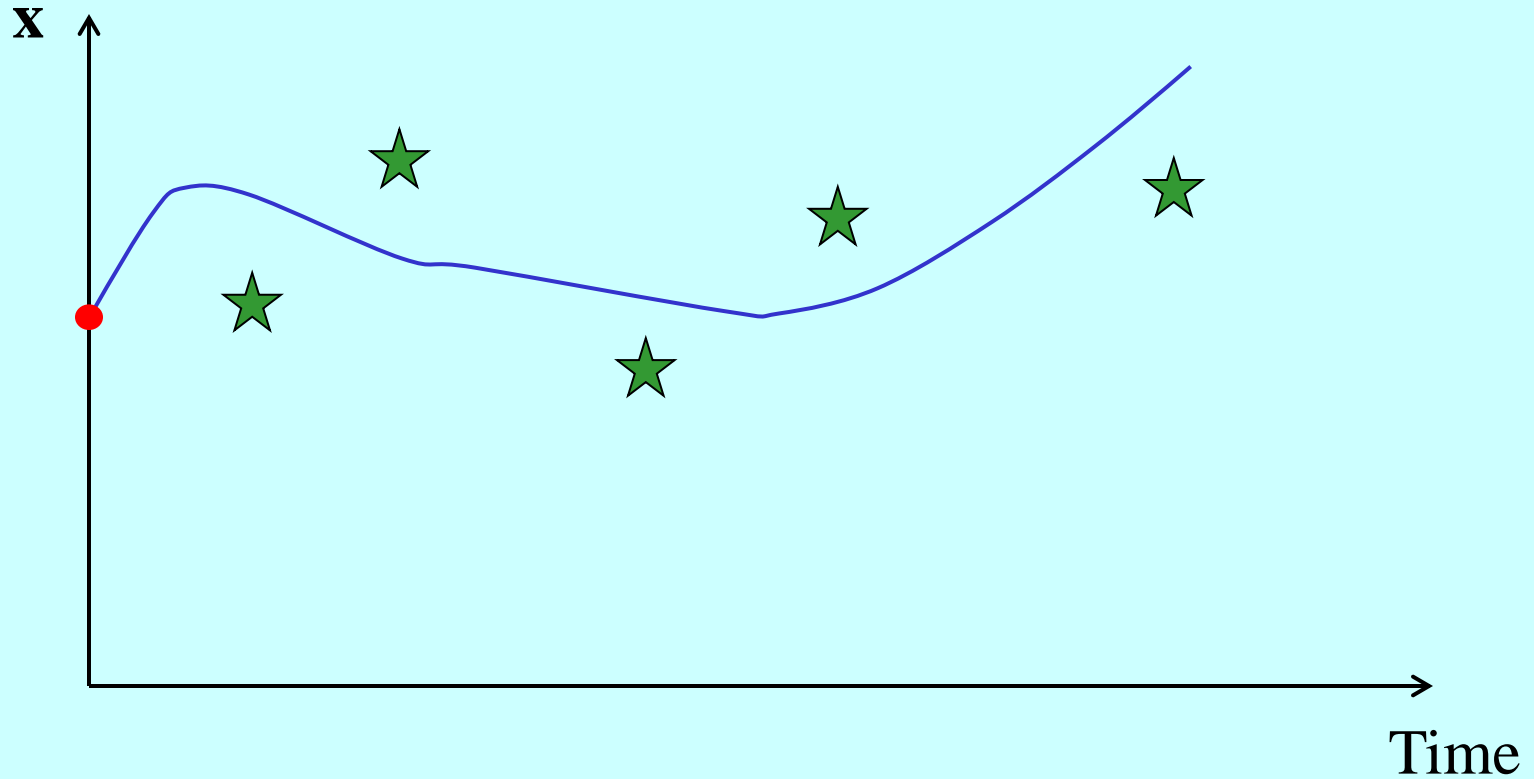
- ★ Observation
- Analysis



# Time sequence of observations

Smoother – Treat all observations together

- ★ Observation
- Analysis



## 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