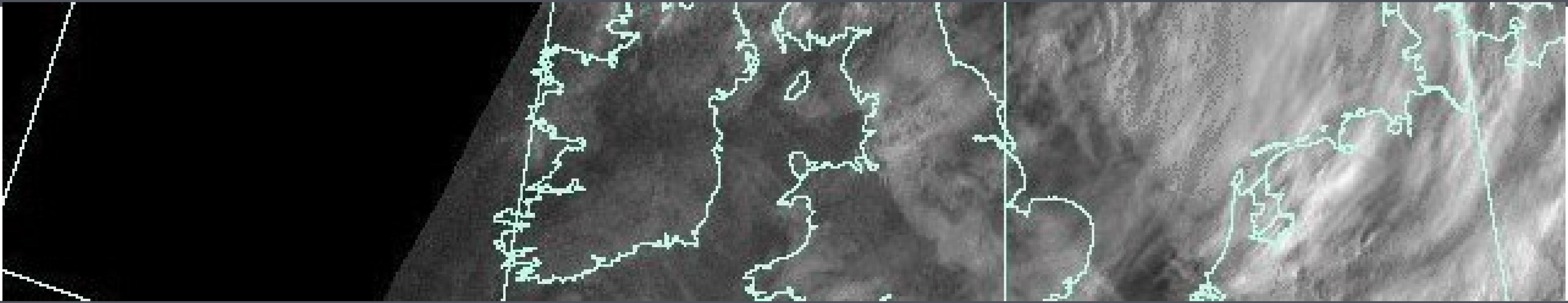
Department of Meteorology
Department of Mathematics & Statistics

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Introduction to Data Assimilation



Prof Sarah L Dance
s.l.dance@reading.ac.uk @DrSarahDance

With thanks to Prof Amos Lawless for a previous version of the slides

Outline

- Why data assimilation?
- What is data assimilation?
- How – is for the rest of the week!

Why data assimilation? Key uses

- **Forecasting** - Using recent observations to improve initial conditions for short-term predictions

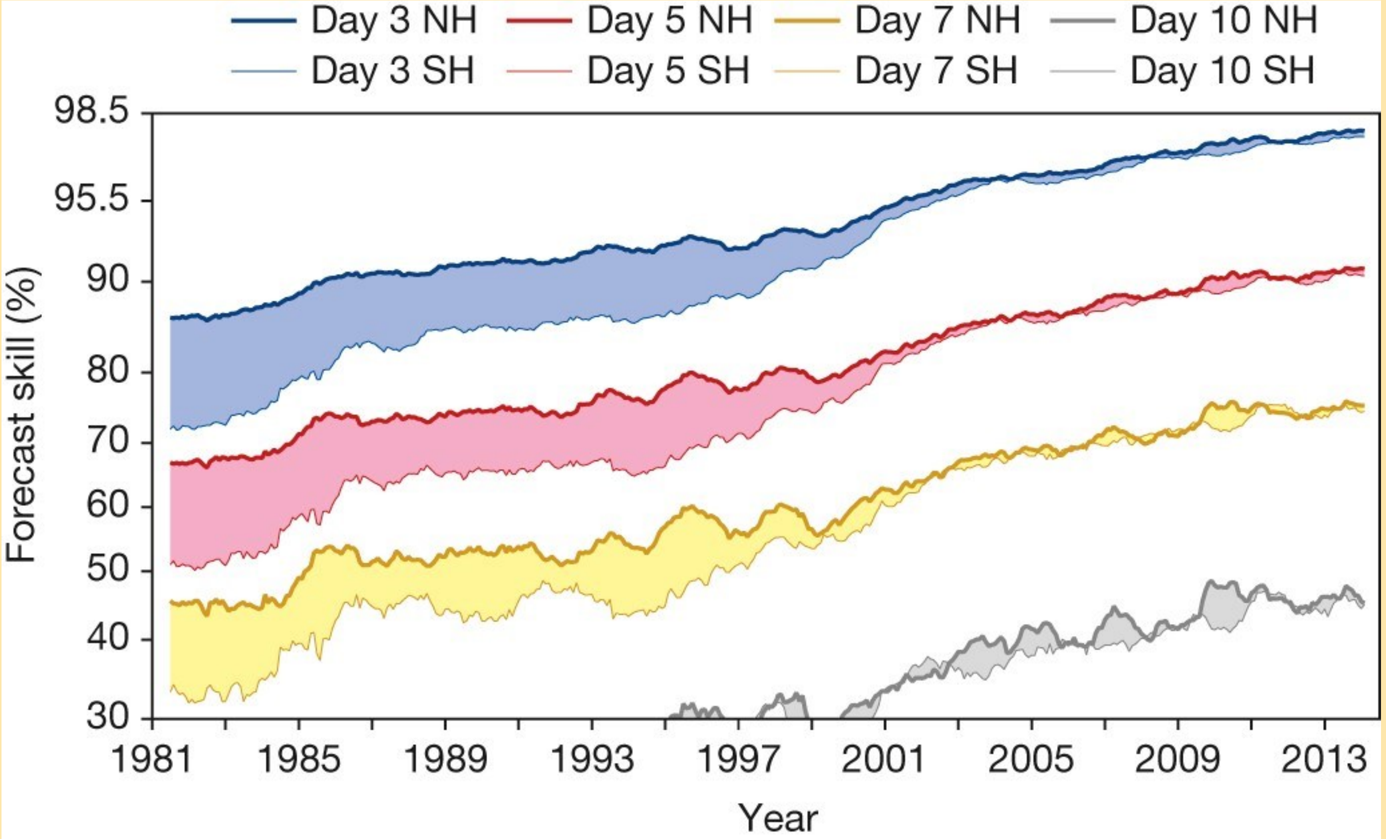
- **Re-analysis:** Learning more about how the Earth works, by using models to interpret/extend different types of data

- **Diagnosis, including parameter estimation:** Testing and improving models by comparing predictions to observations

- **Real-time Control:** Use continually changing estimates of system state to determine control actions

Why data assimilation?

- Initial conditions for a forecast



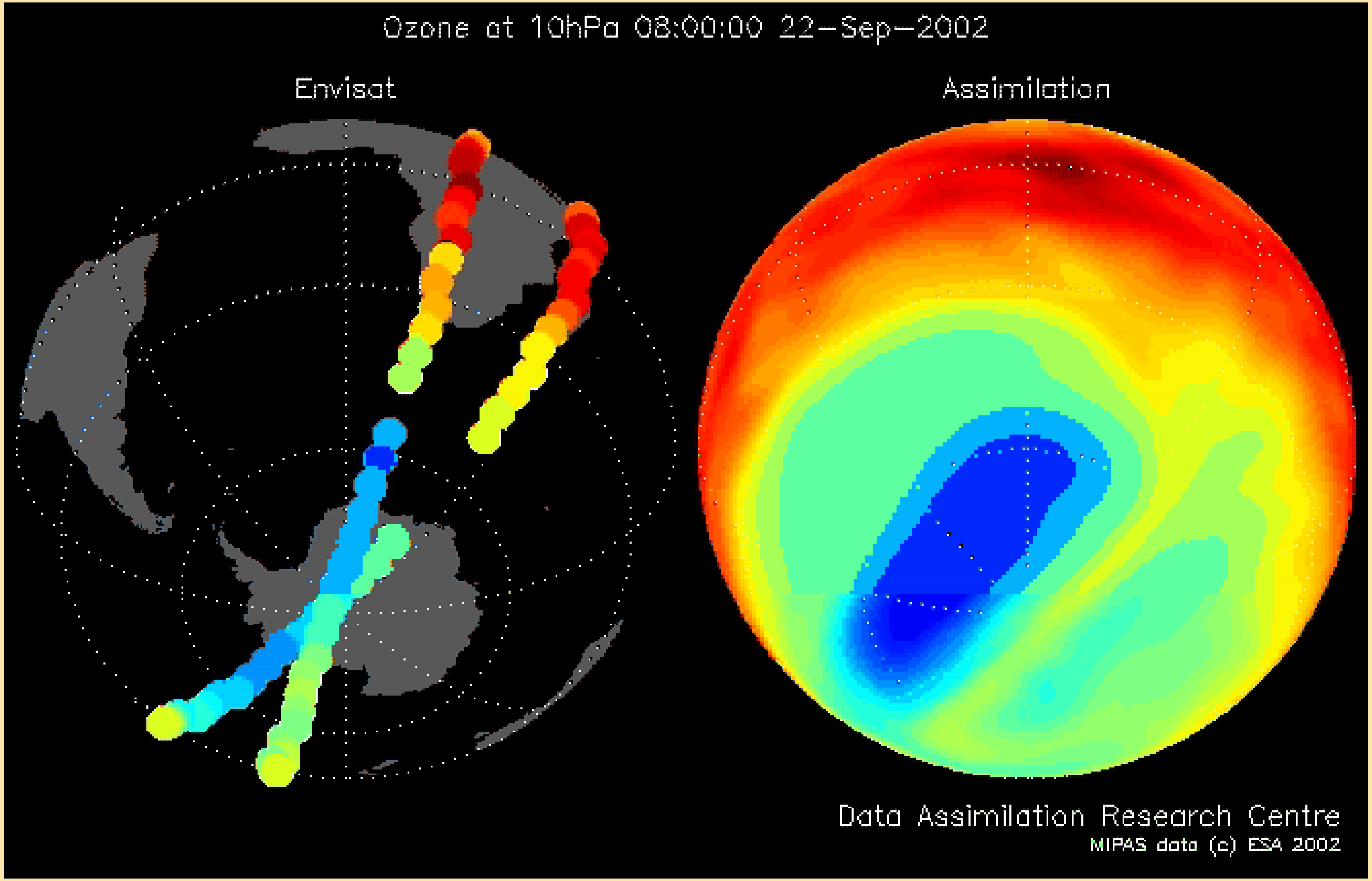
- Example - Steady improvement in global numerical weather prediction skill
- Corresponding improvements in regional forecasts
- In large part due to improved initial conditions for weather forecasts.

Bauer et al. (2005). A measure of forecast skill at 3, 5, 7 and 10-day ranges, computed over the extra-tropical northern and southern hemispheres.

Why data assimilation? Key uses

- **Forecasting** - Using recent observations to improve initial conditions for short-term predictions
- **Re-analysis:** Learning more about how the Earth works, by using models to interpret/extend different types of data
- **Diagnosis, including parameter estimation:** Testing and improving models by comparing predictions to observations
- **Real-time Control:** Use continually changing estimates of system state to determine control actions

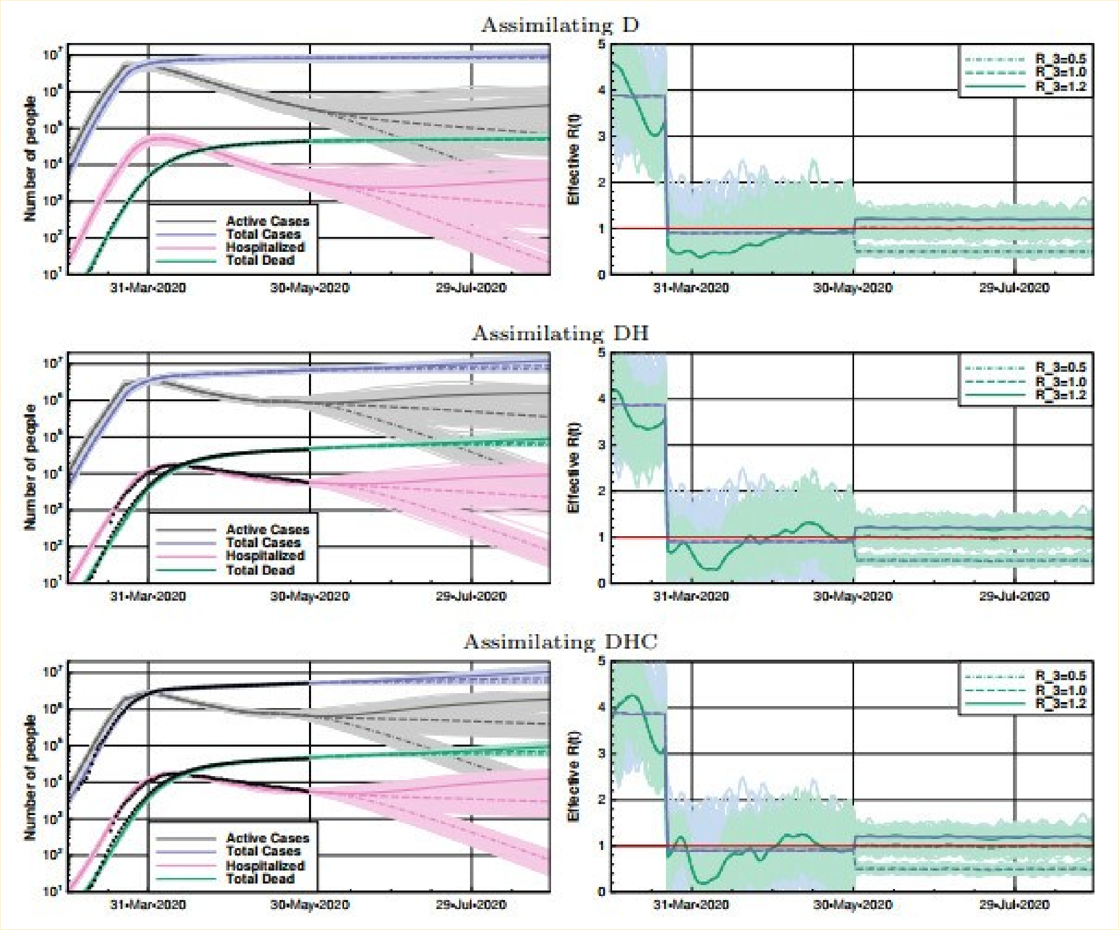
Reanalysis Example – ozone hole



Why data assimilation? Key uses

- **Forecasting** - Using recent observations to improve initial conditions for short-term predictions
- **Re-analysis:** Learning more about how the Earth works, by using models to interpret/extend different types of data
- **Diagnosis, including parameter estimation:** Testing and improving models by comparing predictions to observations
- **Real-time Control:** Use continually changing estimates of system state to determine control actions

Parameter diagnosis example



- Estimation of key parameters during the COVID-19 pandemic using observed data and SEIR models

- Approach was used to provide decision-support in Norway, regarding lockdowns and planning for healthcare resource management (hospital beds, ventilators etc)

Evensen et al (2021) Doi: 10.3934/fods.2021001

Why data assimilation? Key uses

- **Forecasting** - Using recent observations to improve initial conditions for short-term predictions
- **Re-analysis:** Learning more about how the Earth works, by using models to interpret/extend different types of data
- **Diagnosis, including parameter estimation:** Testing and improving models by comparing predictions to observations
- **Real-time Control:** Use continually changing estimates of system state to determine control actions

Real time control example - navigation

- Kalman filters are used in navigation systems (e.g., aircraft, missiles…)
- Need to use good numerical implementations
- Failure of a Patriot Missile to track and intercept an Iraqi Scud missile in Dharan, Saudi Arabia, on February 25, 1991, resulting in the deaths of 28 American soldiers.
- The failure was ultimately attributable to poor handling of numerical rounding errors.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Poll (hands up in the room, poll via teams online)

Which uses of data assimilation are you most interested in? (Choose as many as you like)

https://forms.office.com/e/aH32TBpFzM

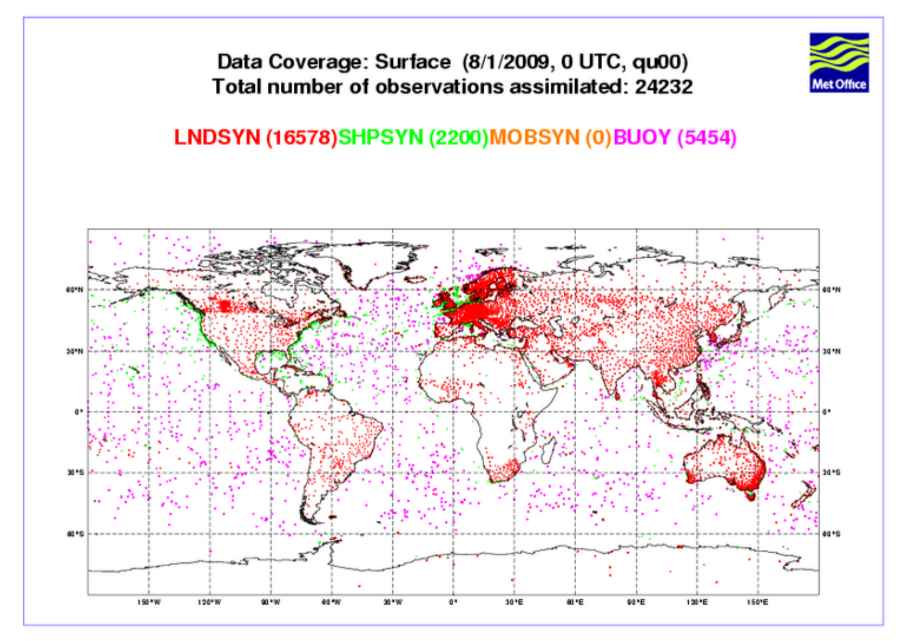
Data Assimilation Introduction Poll

What is data assimilation?

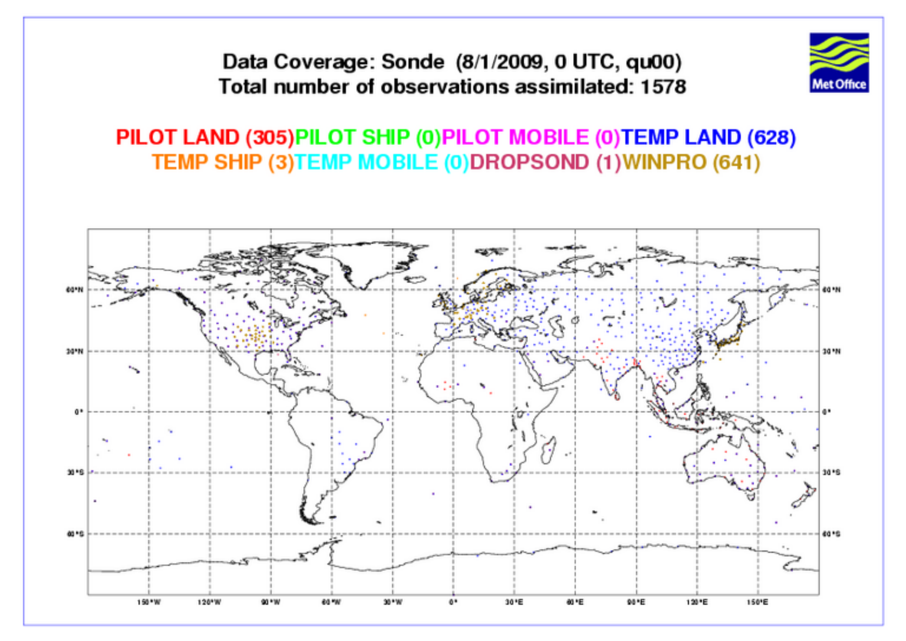
- Data assimilation is the process of estimating the state of a dynamical system by combining observational data with an *a priori* estimate of the state (often from a numerical model forecast).
- We may also make use of other information such as
  - The system dynamics
  - Known physical properties
  - Knowledge of uncertainties

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.

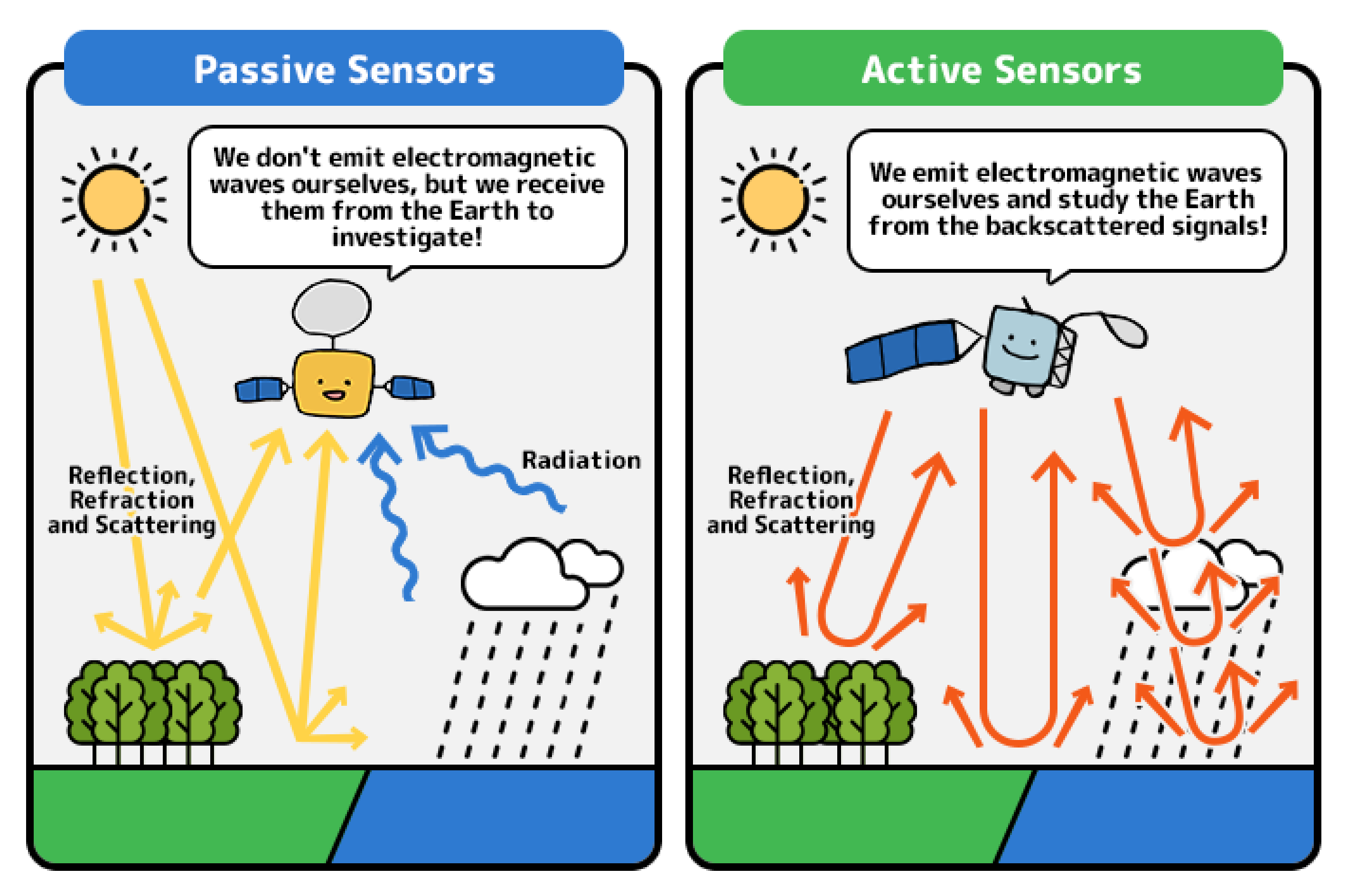
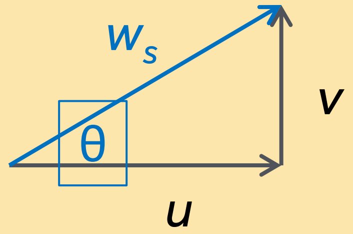


Image from https://www.satnavi.jaxa.jp/en/satellite-knowledge/whats-eosatellite/observation/index.html

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

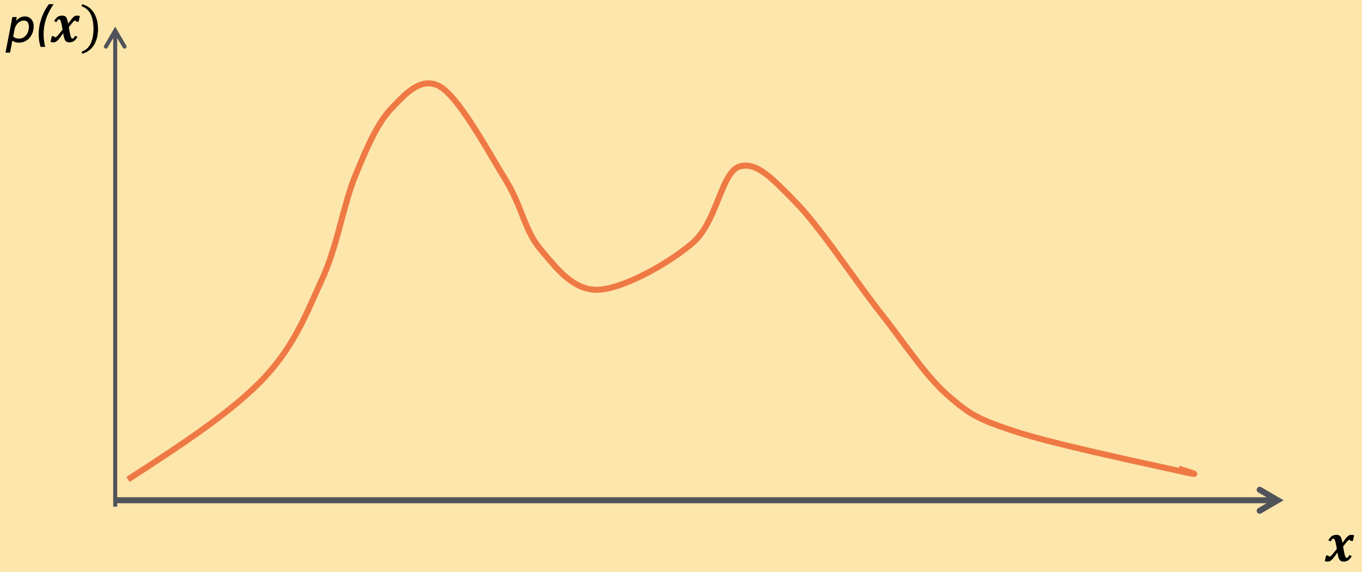
H is known as the observation operator.

Why not just use the observations?

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

Handling the uncertainty

- We need to use probability density functions (pdfs) 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

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

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

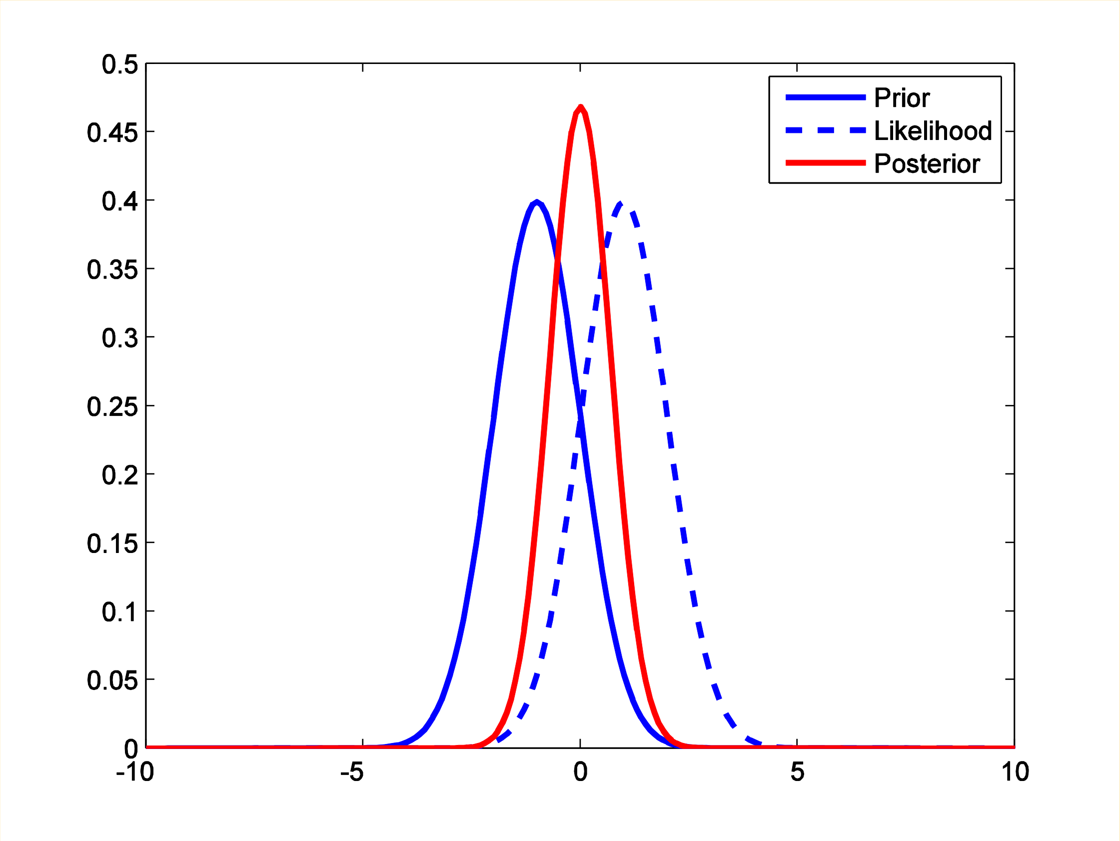
Prior

Likelihood

Posterior

Normalising constant

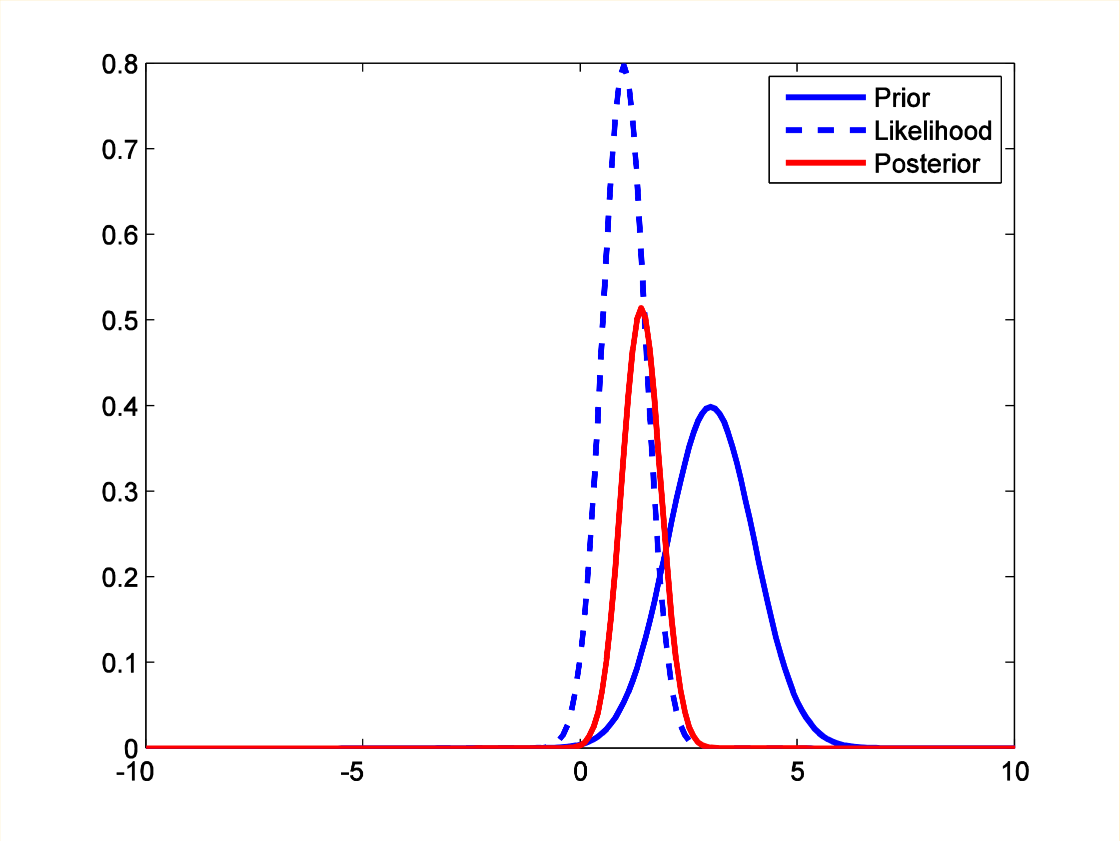
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$



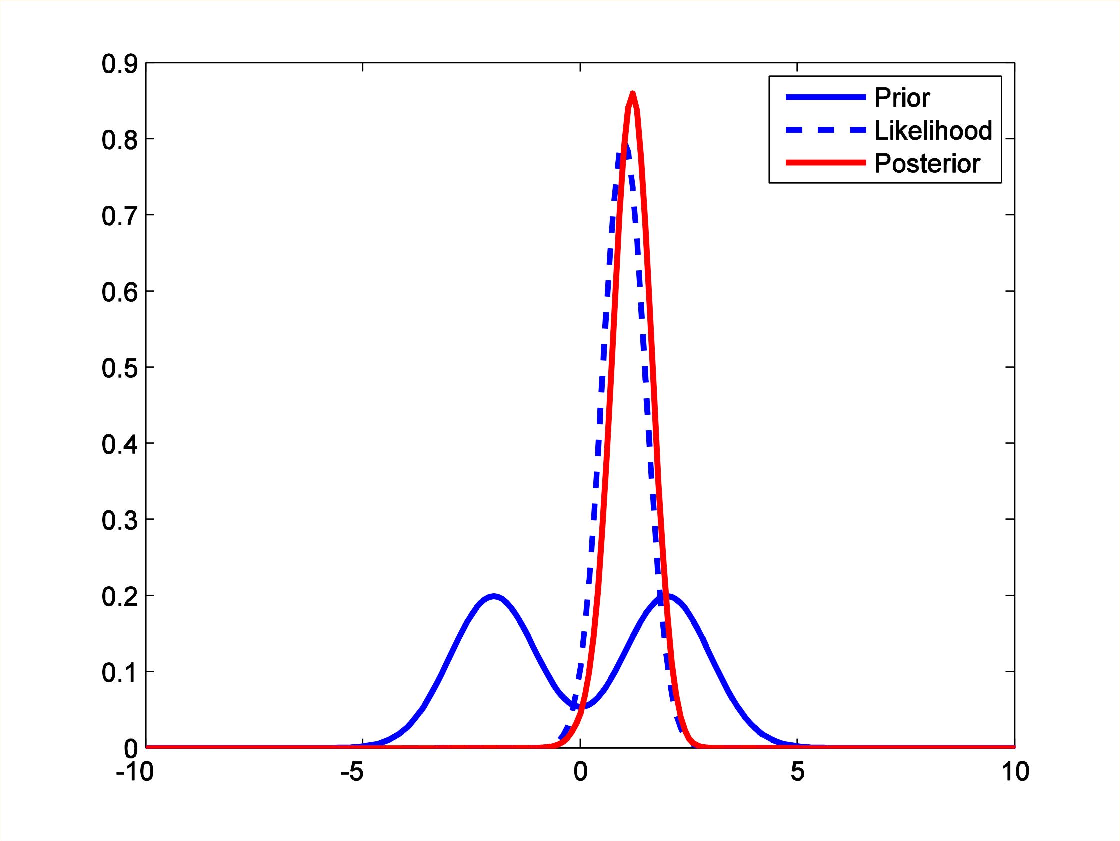
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

Can you explain this plot?

Hint: think about whether you are more confident (less uncertain) in the observations or prior.



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



Practical considerations

In practice the pdfs 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 (aka multivariate normal) 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

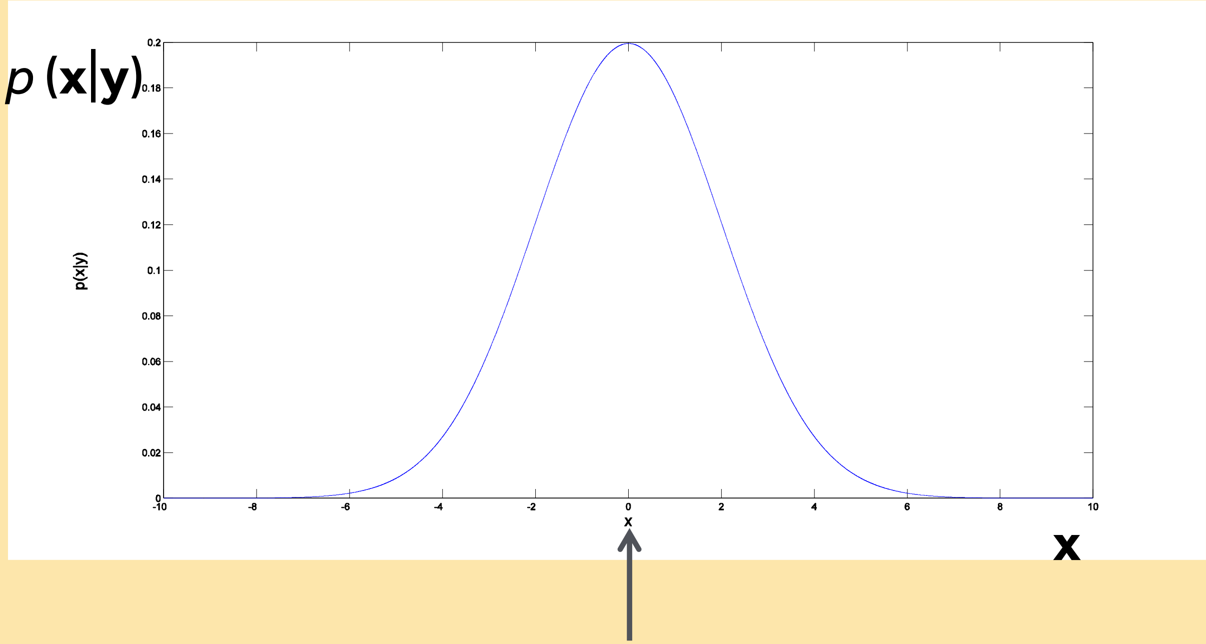
$$p(\mathbf{x}|\mathbf{y}) \propto \exp\{ -\frac{1}{2}\{(\mathbf{x} - \mathbf{x}^b)^T \mathbf{P}^{-1}(\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))\}\}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

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

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

- 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

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

Note size of matrices!

How can we solve this in practice?

1. Variational methods – Directly minimize *J*.
   (Ross Bannister & Amos Lawless - Wednesday)
2. Solve linear equation and approximate covariances with ensemble
   (Ivo Pasmans- Thursday)
3. Hybrid methods – A combination of 1 & 2
   (Ross Bannister - Friday)
4. Beyond the Gaussian assumption – and deep learning
   (Jochen Broecker & Alan Geer – Thursday & Friday)

Summary

- Data assimilation has important uses in forecasting, reanalysis, model diagnosis and real-time control

- Data assimilation provides the best way of using partial observational 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!

Extras

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.

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

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

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

- 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?