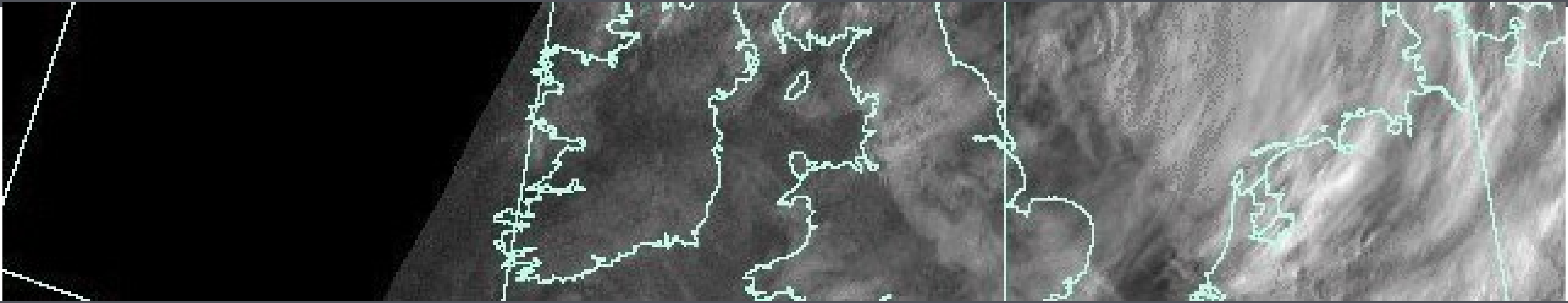
Department of Meteorology
Department of Mathematics & Statistics

Introduction to Data Assimilation



Prof Sarah L Dance

s.l.dance@reading.ac.uk

@DrSarahDance.bsky.social (me)
@unirdg-darc.bsky.social (DARC group)

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

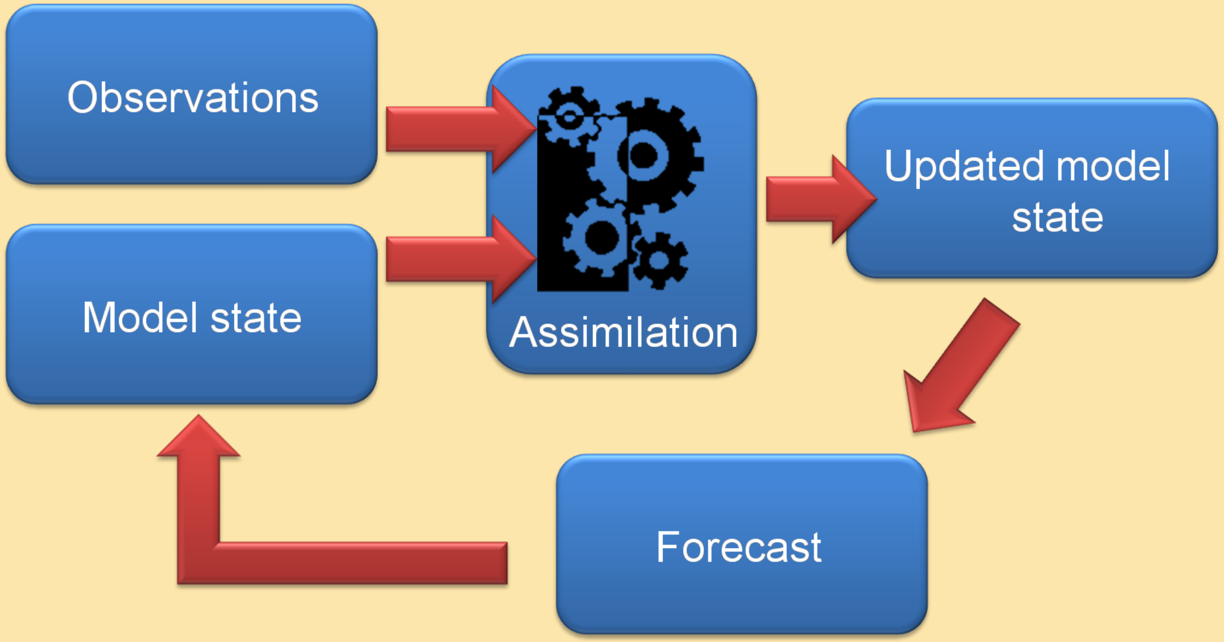
Outline

- Why data assimilation?

- What is data assimilation?

- How – some basic info – more detail in the rest of the week!

Data assimilation



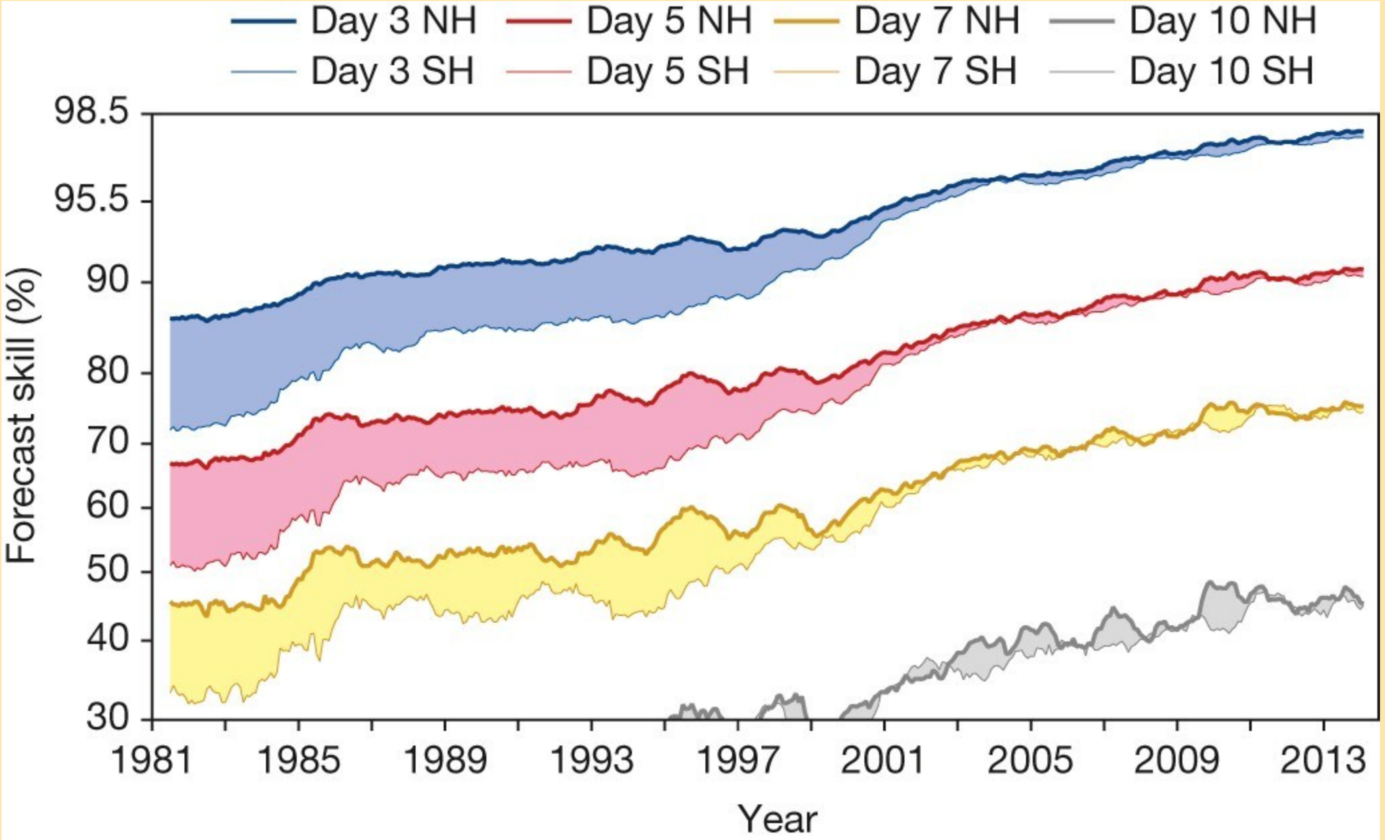
Physics-constrained machine learning technique: updating model information using observations, taking account of uncertainty

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



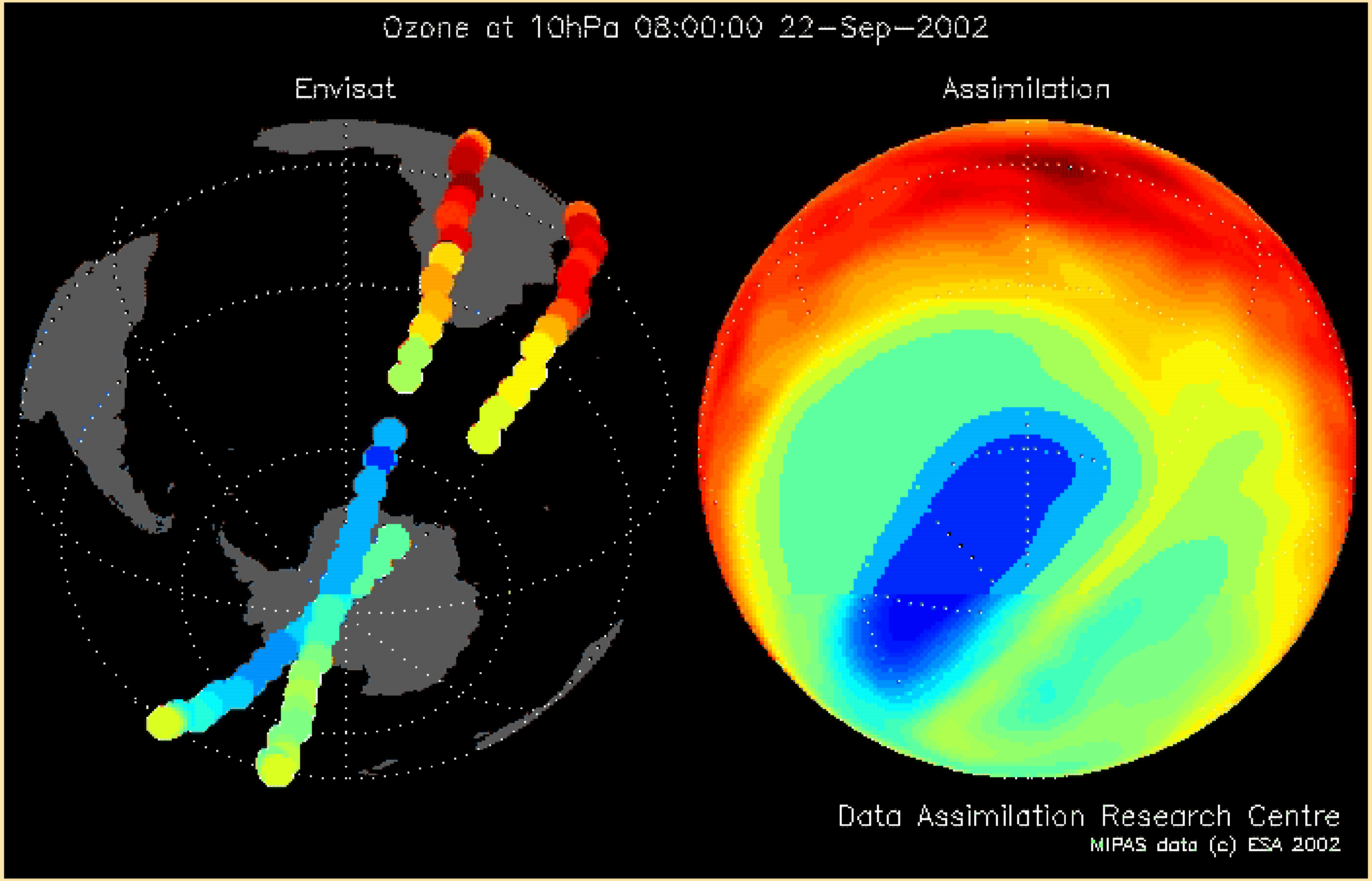
- Example - Steady improvement in global numerical weather prediction skill
- Corresponding improvements in regional forecasts
- In large part due to improved initial conditions for weather forecasts.

Bauer et al. (2015). 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



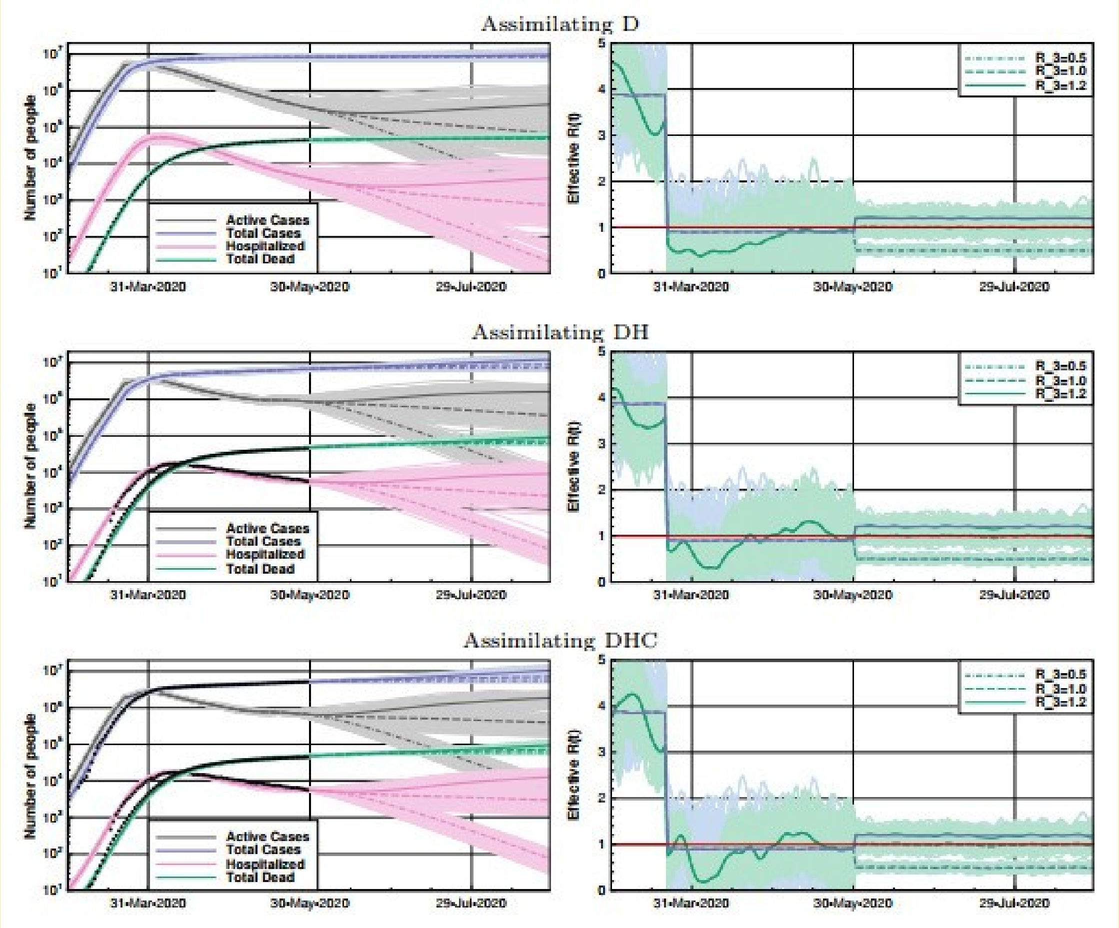
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

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

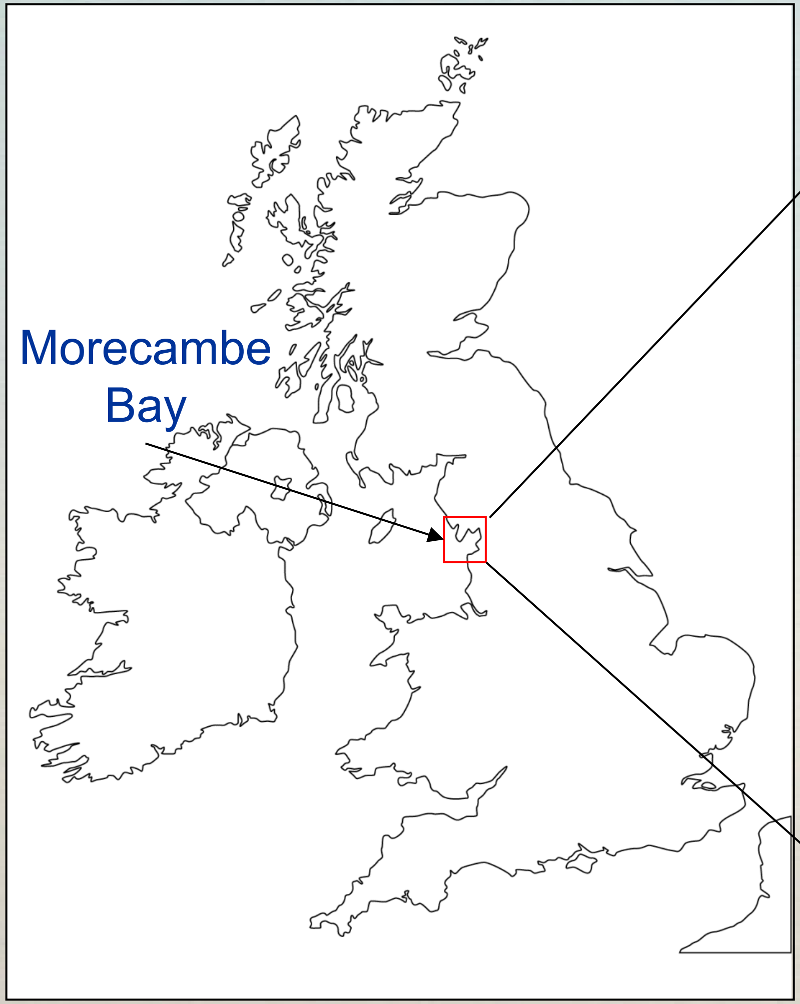
Parameter diagnosis example

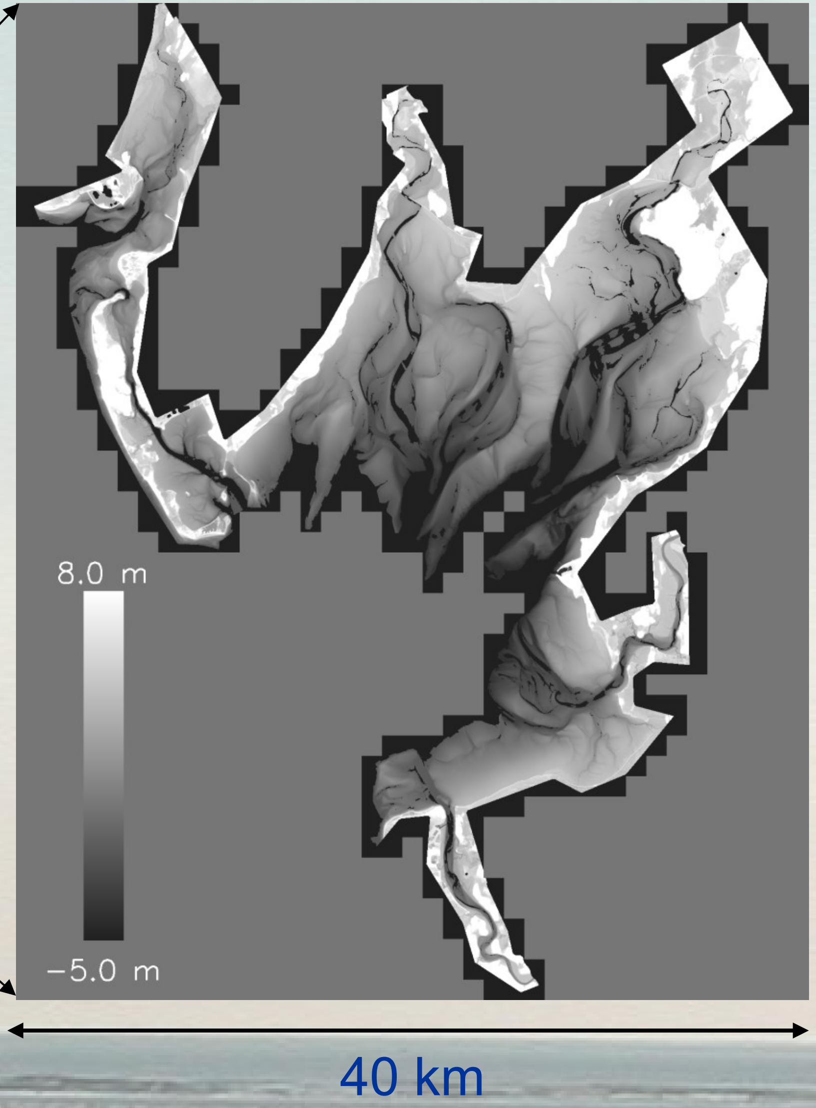


- Estimation of key parameters during the COVID-19 pandemic using observed data and SEIR models
- Approach was used to provide decision-support in Norway, regarding lockdowns and planning for healthcare resource management (hospital beds, ventilators etc)

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

Parameter estimation example





Estuary bathymetry

- Dynamic river channels
- Affects which areas are flooded during a storm-surge event

- Morphodynamic models can predict evolution
- BUT highly uncertain model parameters

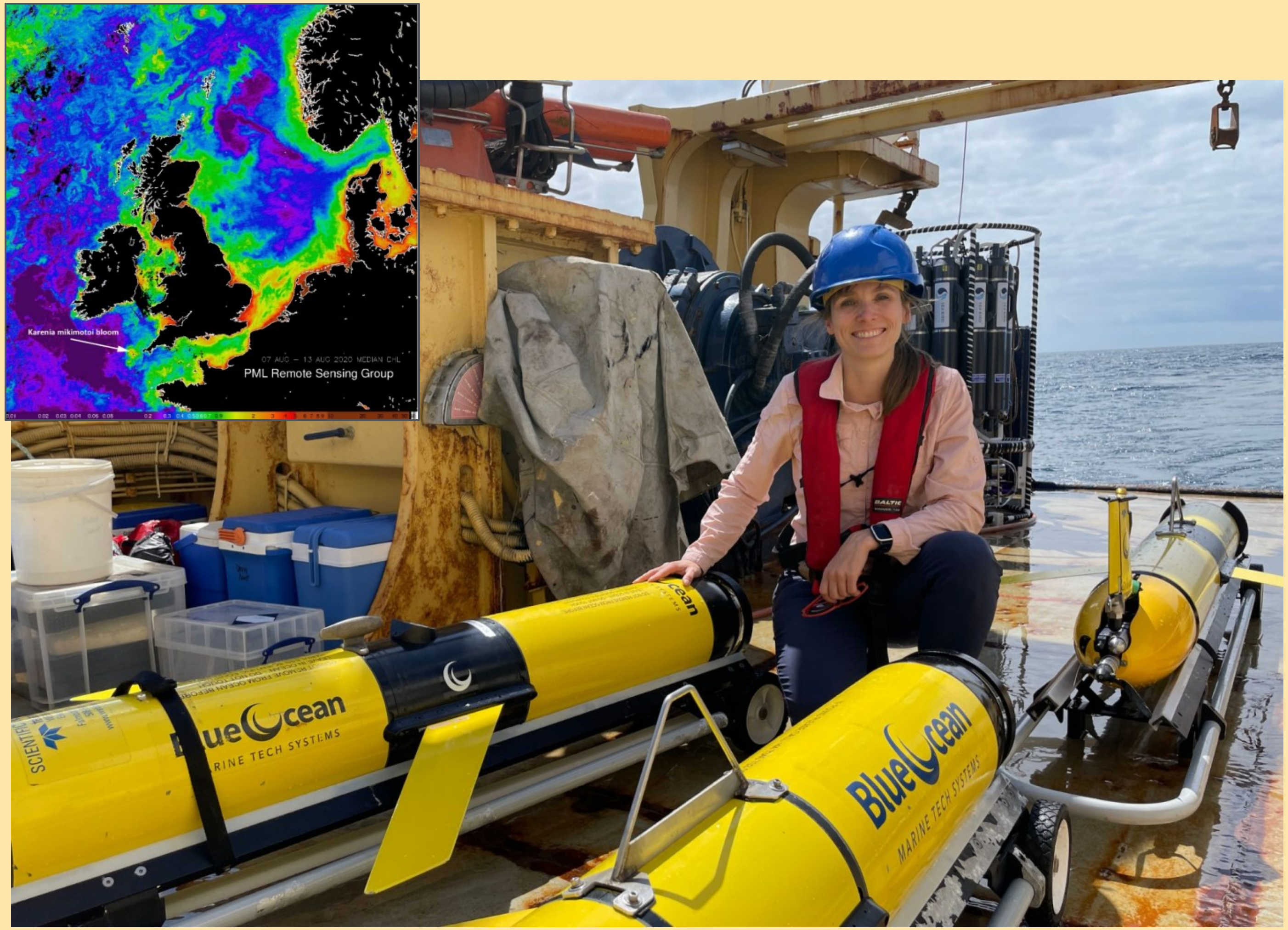
- Use observations to constrain the parameters via data assimilation

e.g. Smith et al (2013) https://doi.org/10.1002/qj.1944

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

Control example – robotic gliders measuring phytoplankton blooms and dissolved oxygen levels



- Robotic ocean gliders from the Plymouth Marine Laboratory
- Measuring phytoplankton blooms and dissolved oxygen levels
- Data assimilation with a coupled shelf-sea biogeochemistry model
- Informs gliders where to go next to maximize value of observations for predictions

https://pml.ac.uk/news/pml-successfully-deploy-a-fleet-of-ocean-robots-to/

Poll

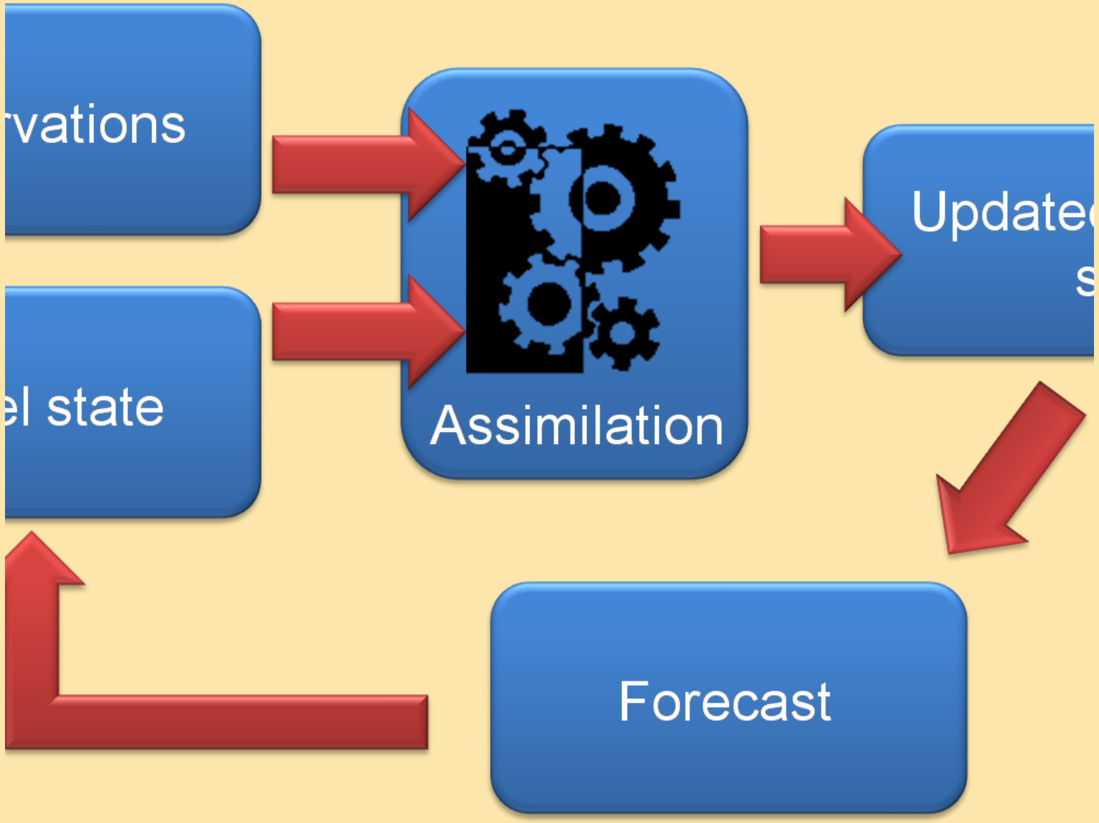
Which uses of data assimilation are you most interested in? (Choose as many as you like)

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

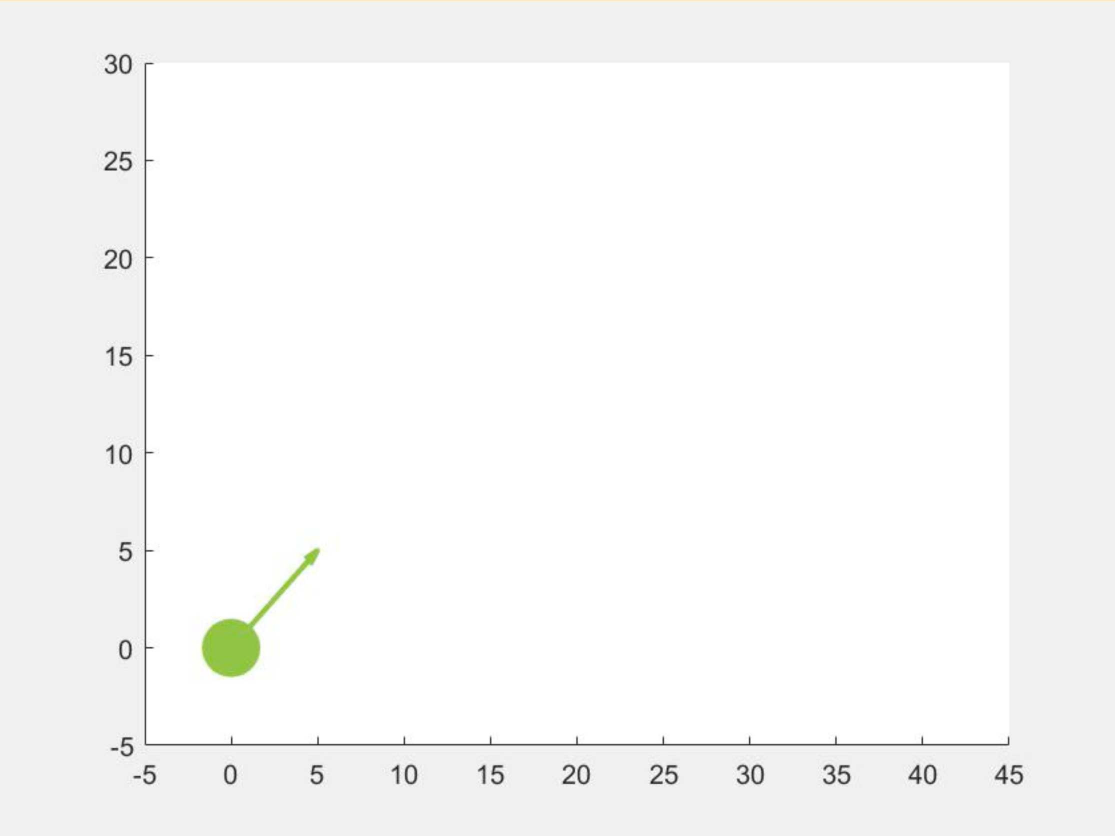


Dynamical model

- A model is a simplified representation of the real world
- In earth-system-modelling we usually solve mathematical equations describing a physical (biological, chemical) process as a function of time
- This enables us to make predictions of the future, or study events from the past
- “Data driven” models that are developed through deep learning are becoming more popular, but they still need to be initialized.

Data assimilation is often thought of as a way of keeping a model “on track” by constantly correcting it with fresh observations

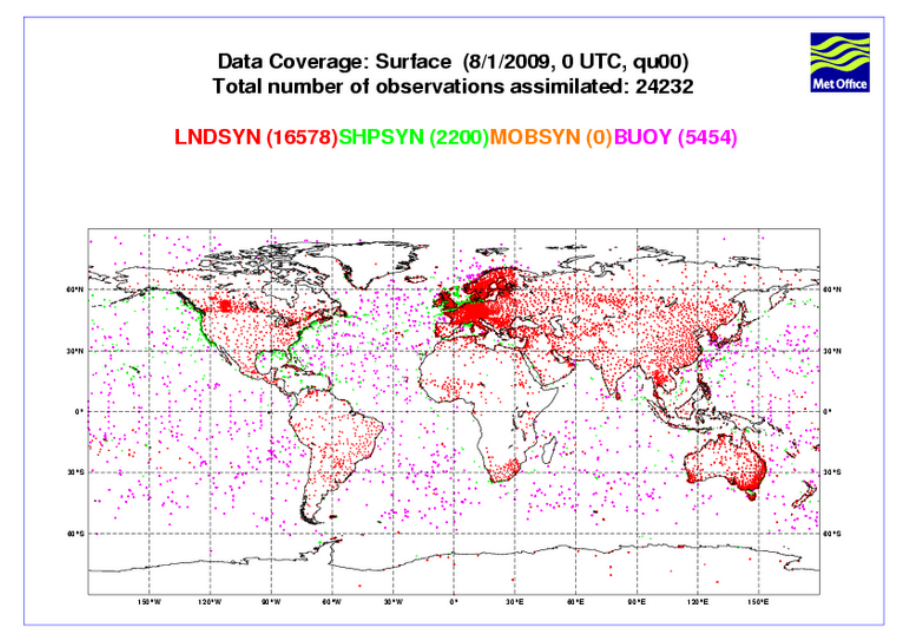
Initial value problems



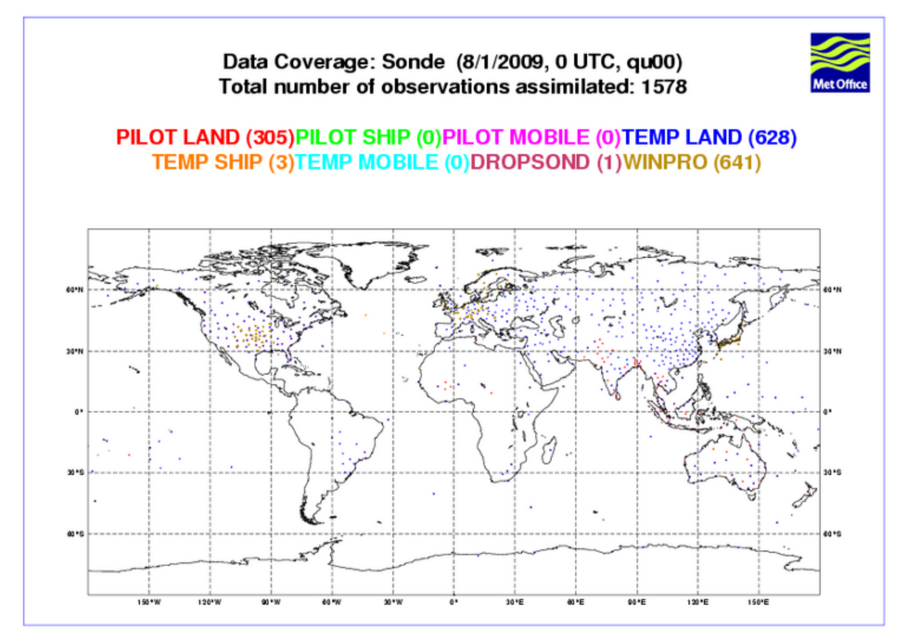
- In data assimilation we often estimate the initial conditions for a model prediction
- Often the models are described by (partial) differential equations discretized in a computer
- We have an “initial value problem”

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.

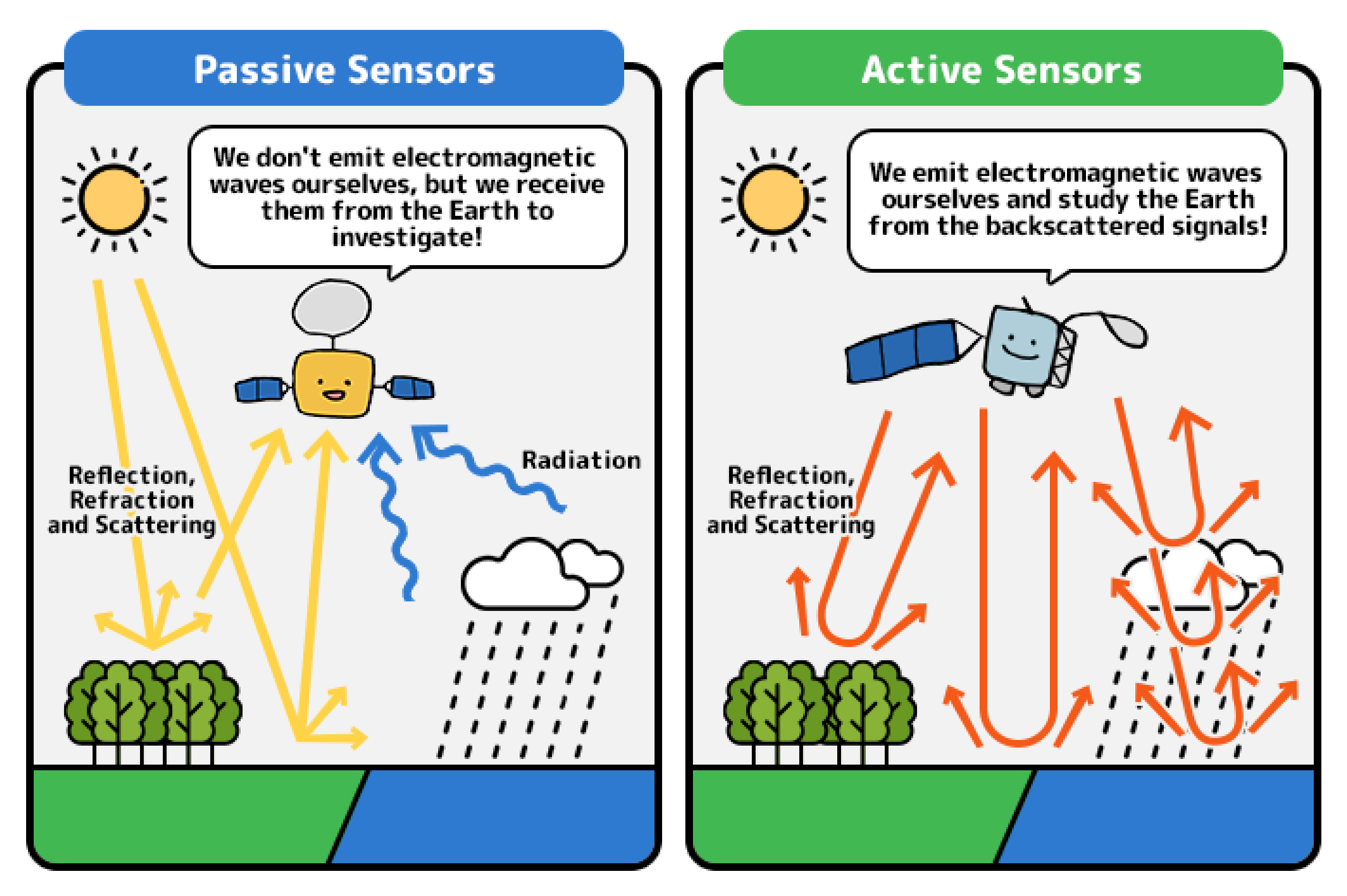
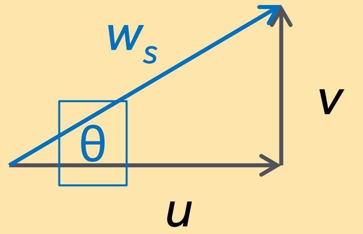


Image from https://www.satnavi.jaxa.jp/en/satellite-knowledge/whats-eosatellite/observation/index.html

Example

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



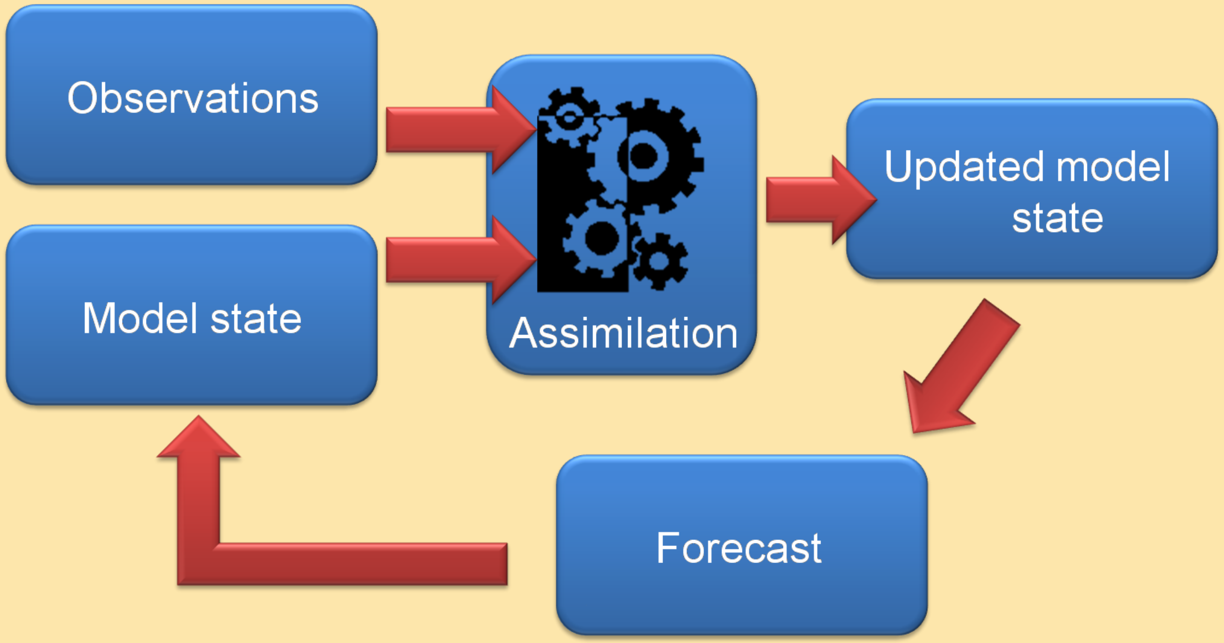
Suppose we observe the wind speed w_s.

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

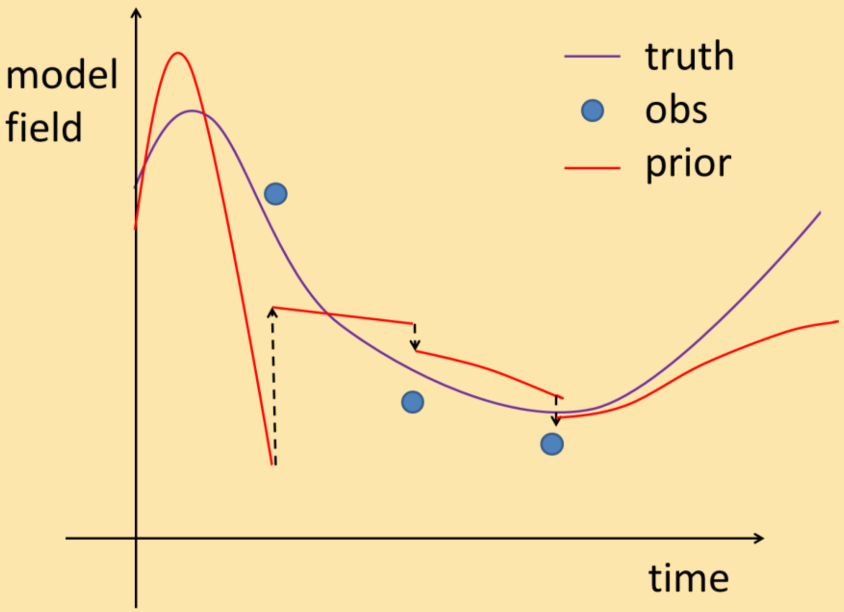
with $H(\mathbf{x}) = \sqrt{u^2 + v^2}$

H is known as the observation operator.

How do we do data assimilation ?



Example - 3D-Variational Data Assimilation (3D-Var)

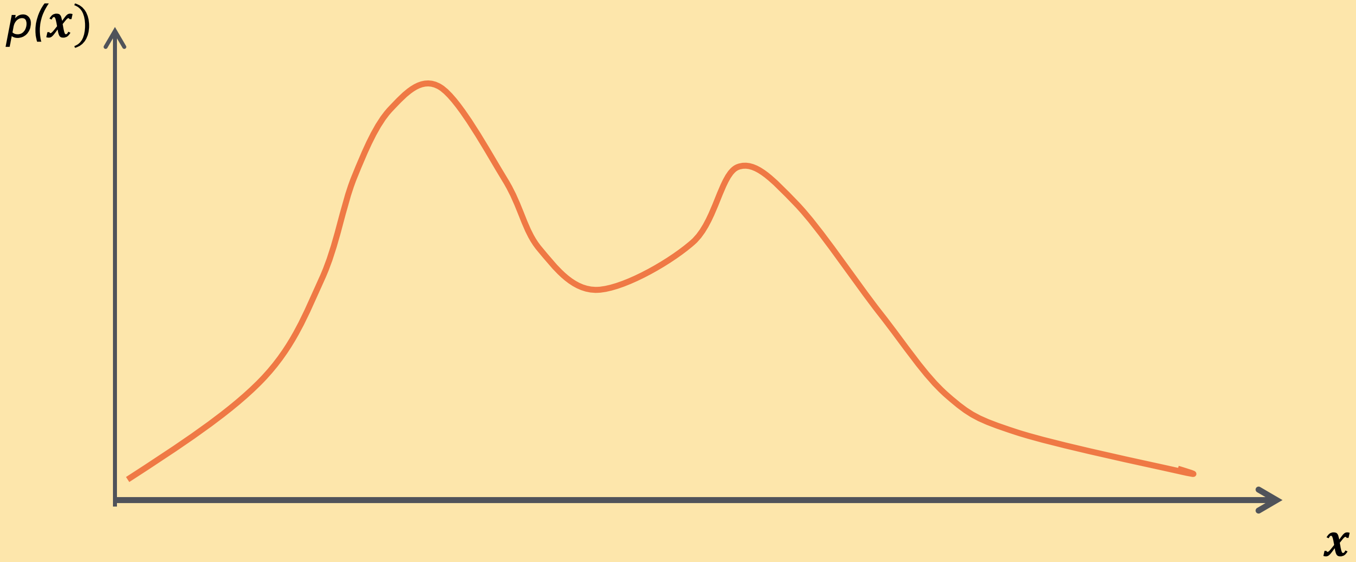


Minimize

$$J(\mathbf{x}) = \left(\mathbf{x} - \mathbf{x}^b\right)^T \mathbf{B}^{-1} \left(\mathbf{x} - \mathbf{x}^b\right) + (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}) \,.$$

Handling the uncertainty

- We need to use probability density functions (pdfs) to represent the uncertainty and give the correct weight to the observations and prior model state.



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

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

Prior

Likelihood

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

Posterior

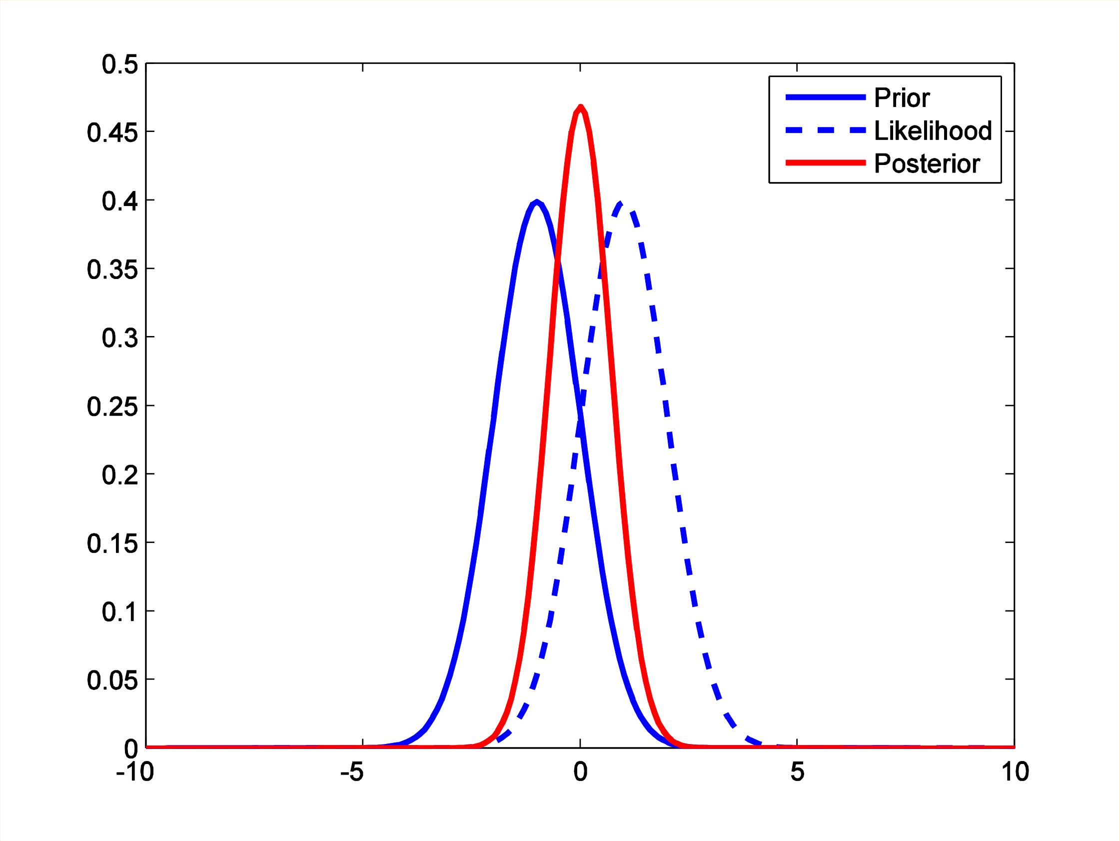
Normalising constant

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

Example

- Model temperature prediction (prior)
  -2^{O}C ± 2
- Observed temperature (likelihood)
  +2^{O}C ± 2

- Equal weight to observation and prior

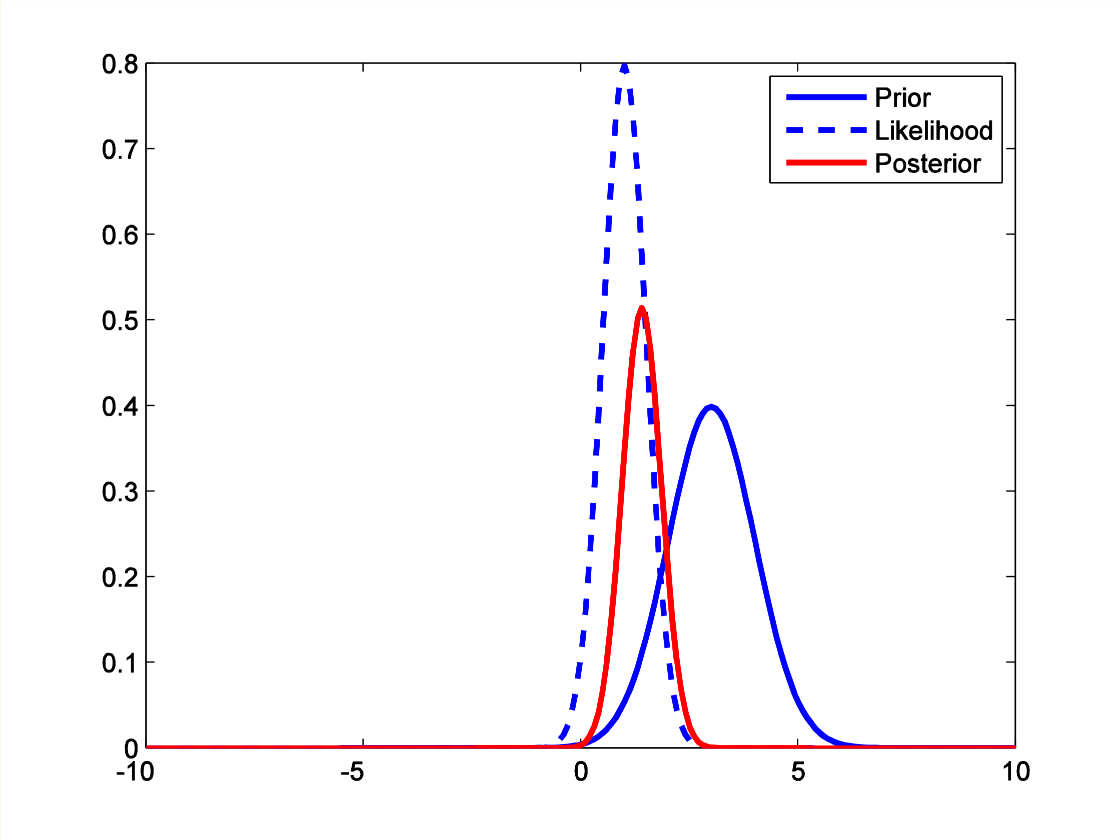


$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

Pause

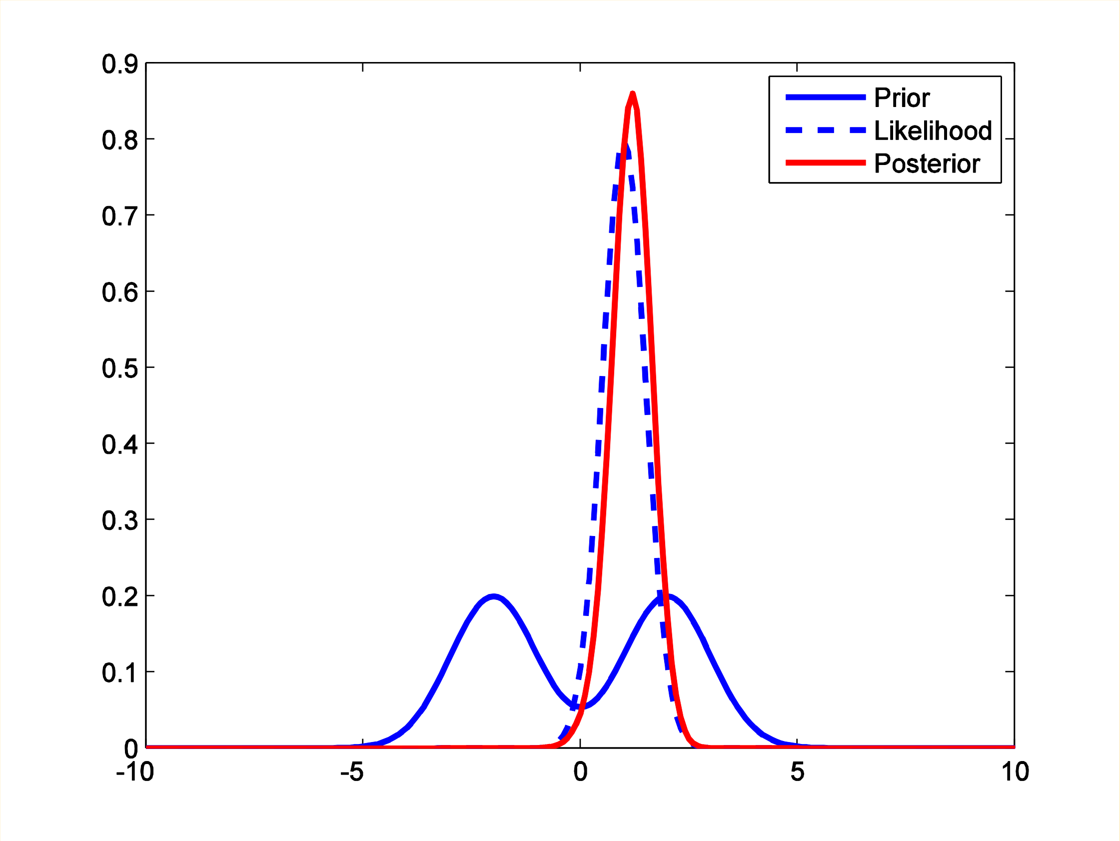
Can you explain this plot?

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

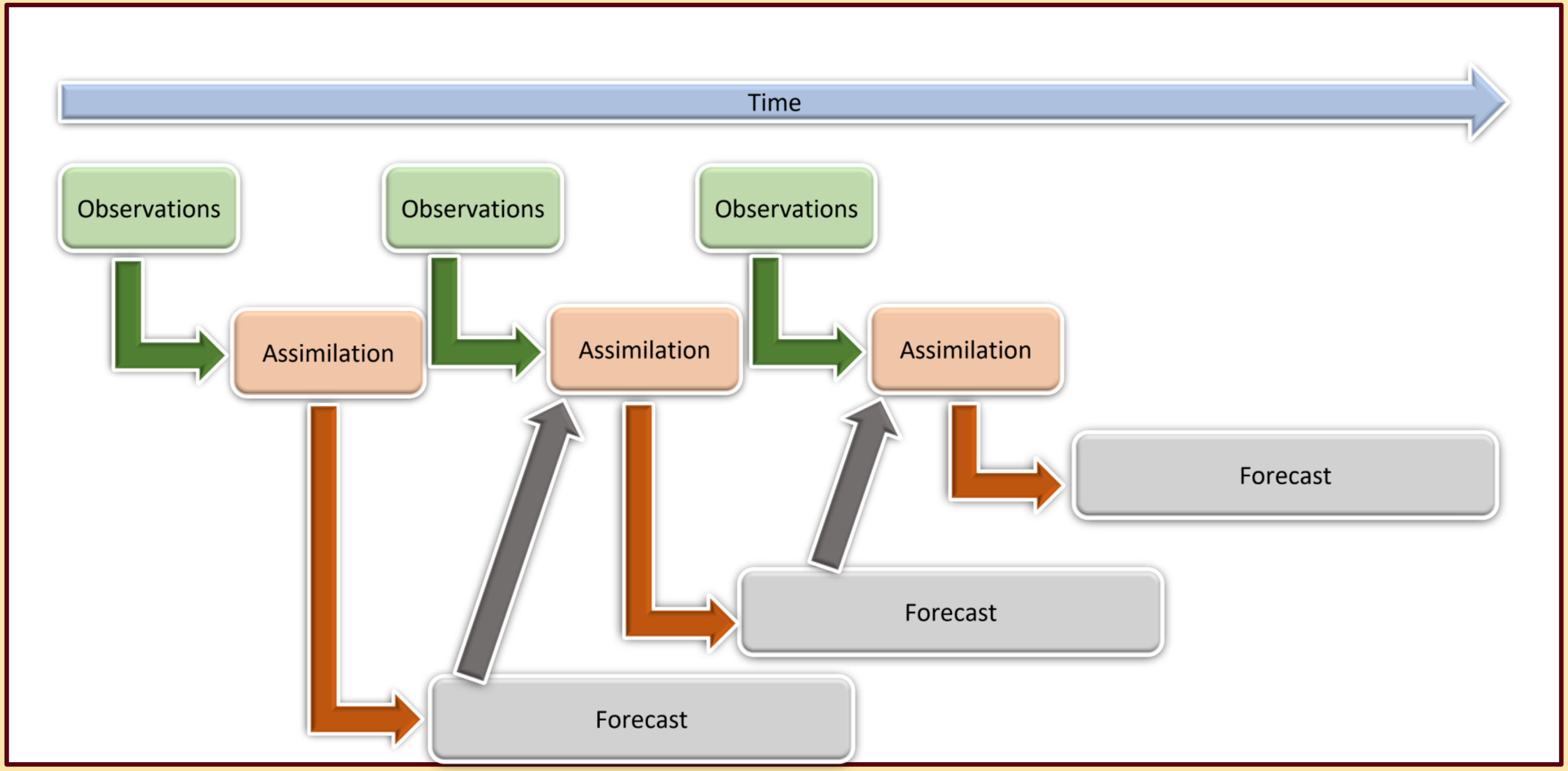


$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

Not all probability distributions are Gaussian



Practical considerations



In practice the pdfs may be very high dimensional (e.g. 10^9 in Numerical Weather Prediction).

Time constrained (both start and finish)

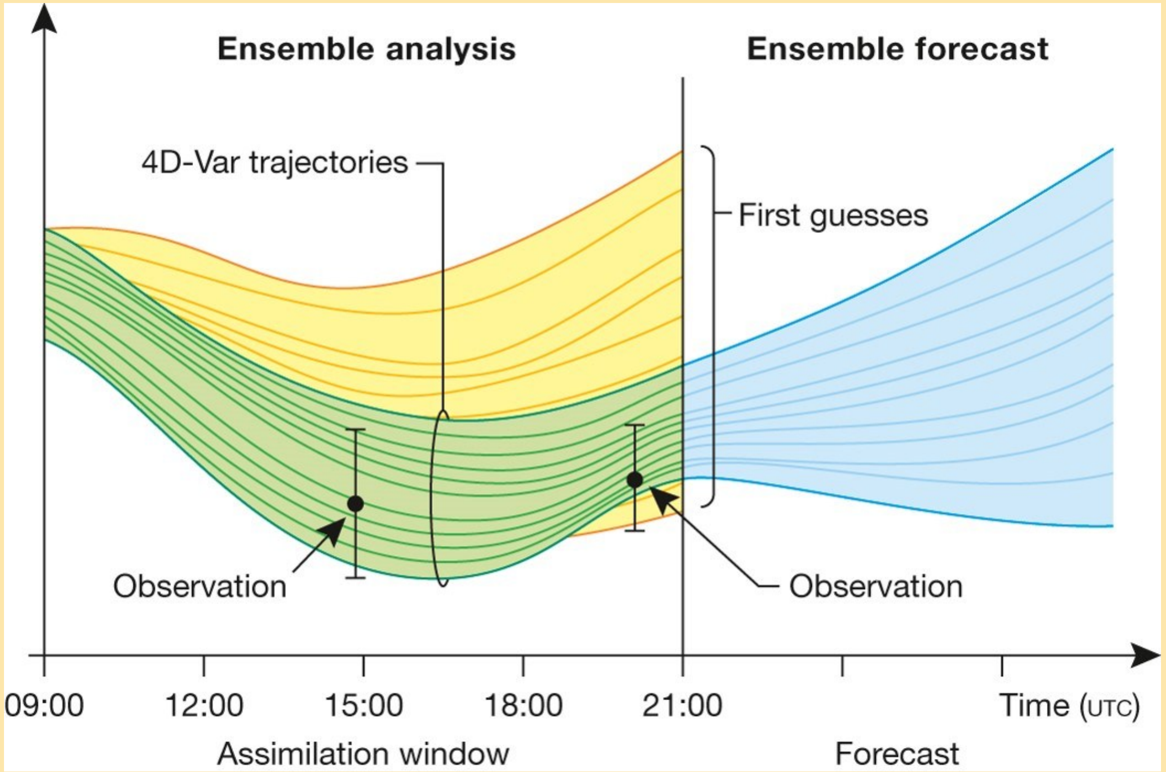
Practical considerations

This means

- We cannot calculate the full pdf.

- Consider using numerical approximations and/or machine learning techniques to accelerate our computations

- E.g. ensemble estimates of the prior



Bauer et al. (2015). https://doi.org/10.1038/nature14956

Summary and knowledge check

- 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!

Introduction to data assimilation knowledge check quiz

https://forms.office.com/e/hCJnH7Un2v

View results to see feedback