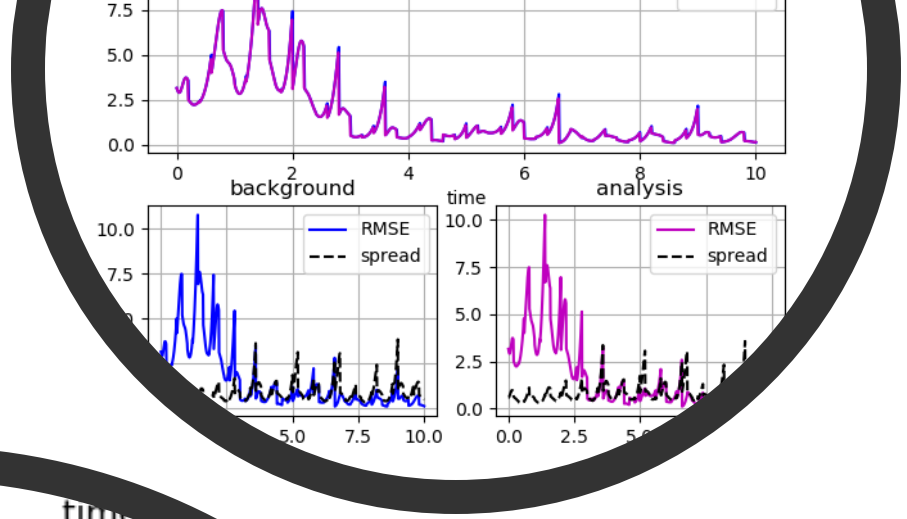
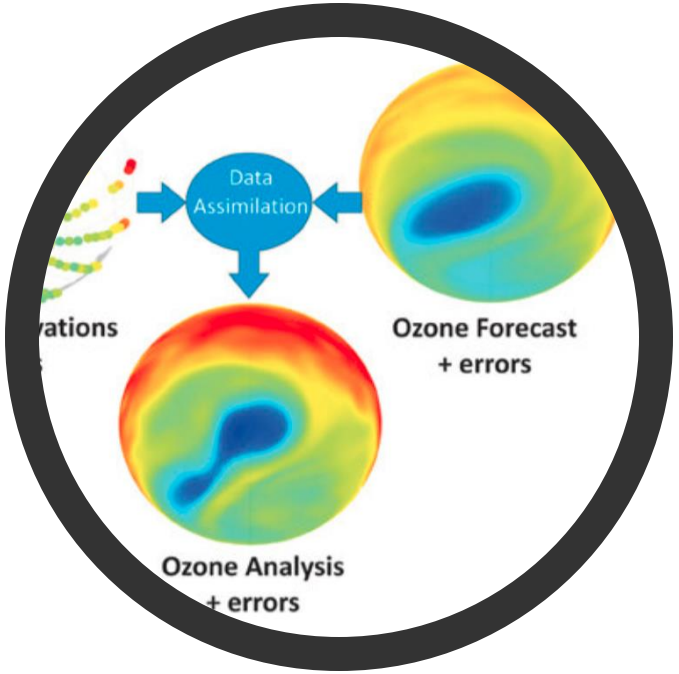
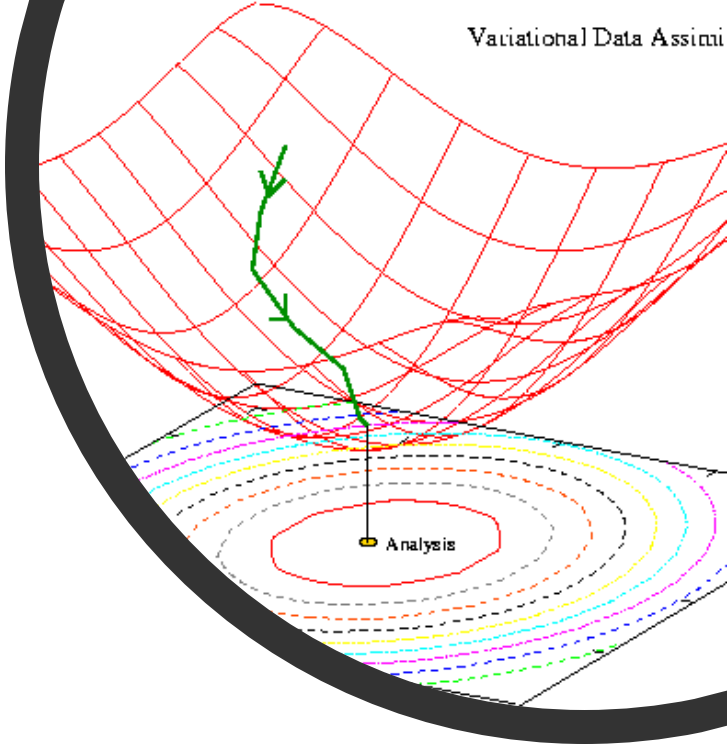
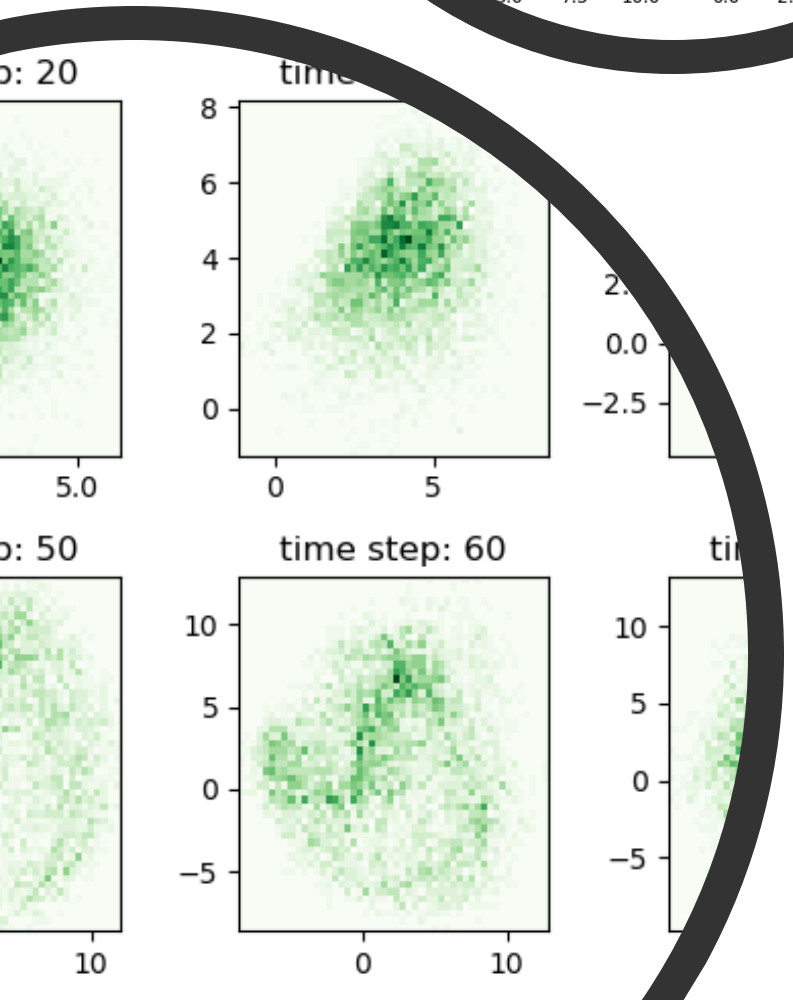


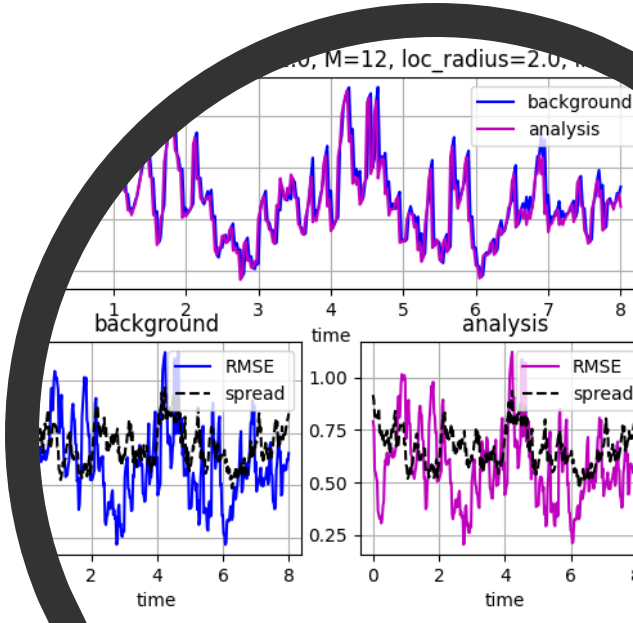






Alison Fowler

Overview of DA algorithms



NERC/NCEO/DARC Training course on data assimilation
and its interface with machine learning 2025

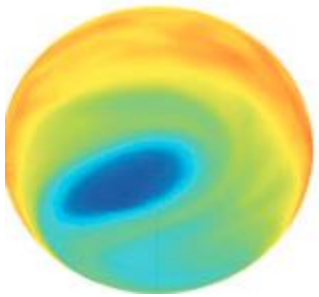
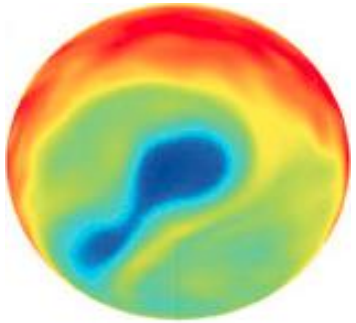
Aim of lecture

- Give an overview of the different approaches to solving the data assimilation problem by building on a common framework.
- Introduce the frequently used terminology
- Highlight the similarities and differences between the different approaches.

- More detailed descriptions of the algorithms will be presented in the rest of the week.

Recap aims of DA

The DA problem is to combine **prior knowledge** and **relevant observations** to give an **updated estimate of the state** of the atmosphere/oceans/land surface etc.

DA( ) = 

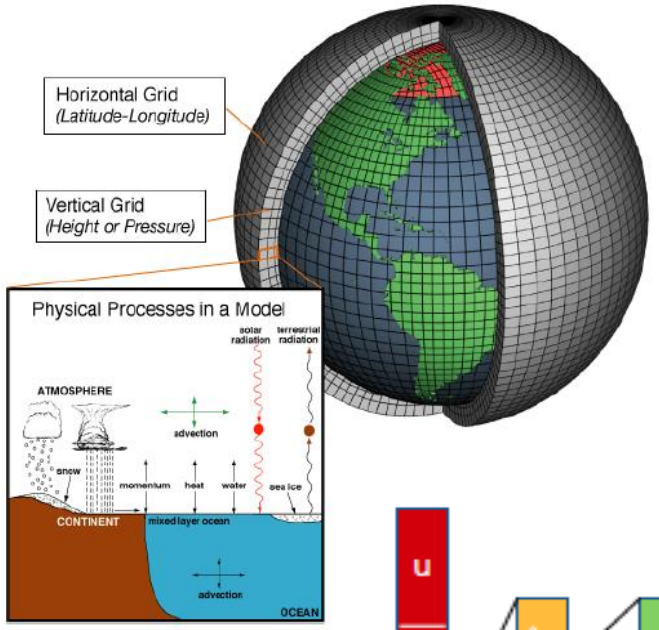
The prior e.g. discretised model and uncertainties

Observations and uncertainties

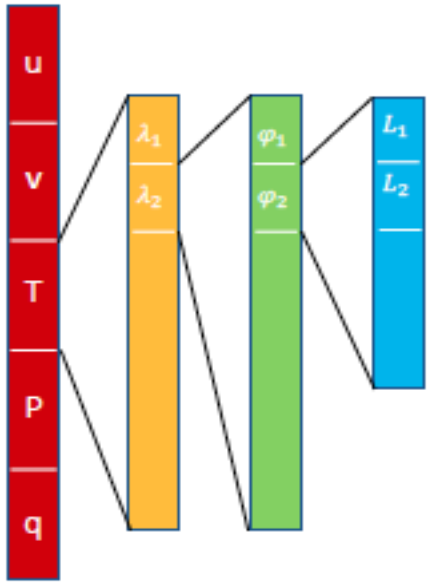
Updated estimate e.g. discretised analysis

- To **initialize a forecast** – the better the initial conditions the better the forecast.
- Create **reanalyses** to understand the recent past.
- **Estimate parameters** in the model to give a better understanding of the processes represented.

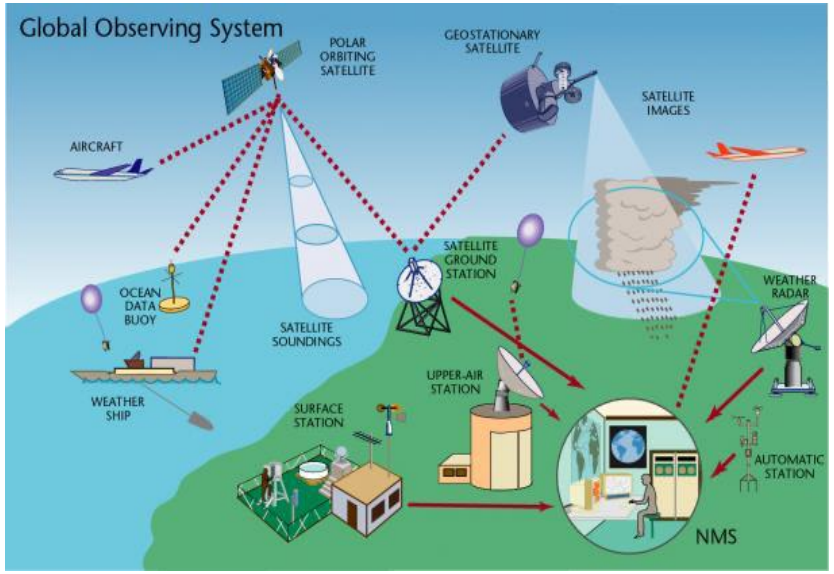
Linking observations to models



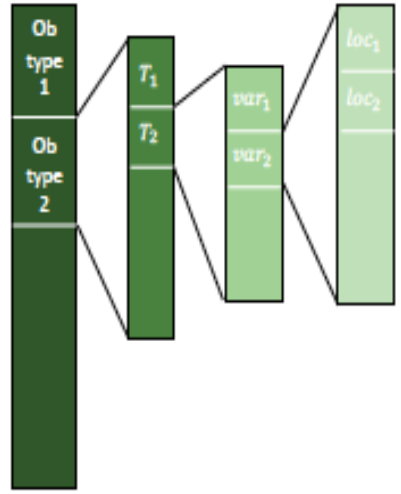
$\mathbf{x} \in \mathbb{R}^n$ is a vector of the model variables that we want to estimate. Referred to as the *STATE*.



As $\mathbf{x}$ and $\mathbf{y}$ lie in different spaces it is necessary to define an *OBSERVATION OPERATOR*, h: $\mathbb{R}^n \longrightarrow \mathbb{R}^p$, that maps $\mathbf{x}$ from state space to observation space.



$\mathbf{y} \in \mathbb{R}^p$ is a vector of the available observations



The inverse problem

The observation operator, which maps from state to observation space, is often known as the *FORWARD MODEL*.

Reconstructing $\mathbf{x}$ from $\mathbf{y}$ is often referred to as the *INVERSE PROBLEM*.

If the observation operator were linear, we could write $\mathbf{y} = \mathbf{H}\mathbf{x}$, where $\mathbf{H} \in \mathbb{R}^{p \times n}$.

Usually $\mathbf{H}$ is non-square ($n \neq p$) and/or rank deficient. Therefore, we cannot invert $\mathbf{H}$ to find $\mathbf{x} = \mathbf{H}^{-1}\boldsymbol{y}$ directly.

Additionally, the observations are not perfect and contain errors.

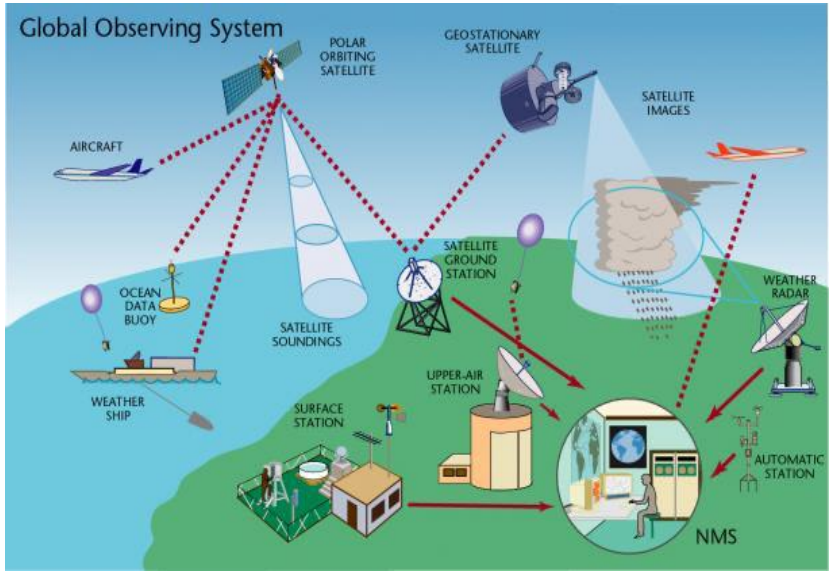
Data assimilation provides a framework for solving the inverse problem by introducing prior information about $\mathbf{x}$.

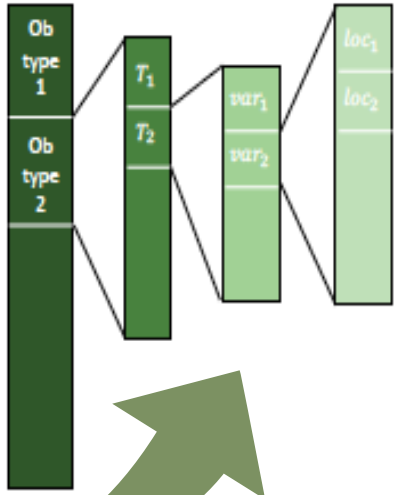
Bayes' theorem

Most DA algorithms can be derived from Bayes' theorem:

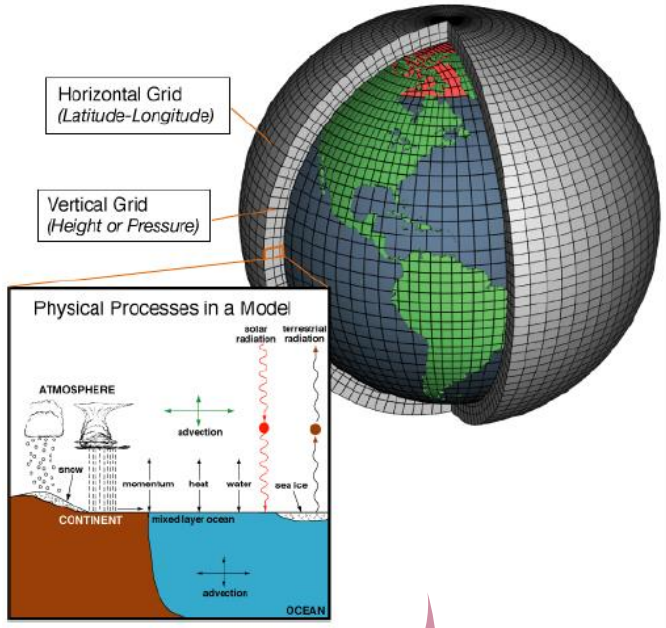
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

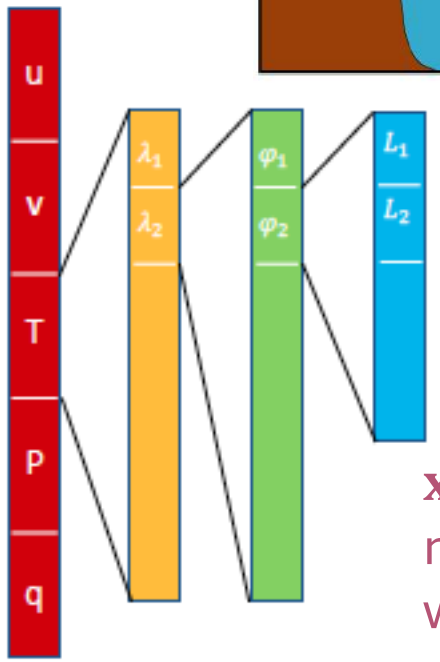
posterior; prior; likelihood; marginal





y is a vector of the available observations





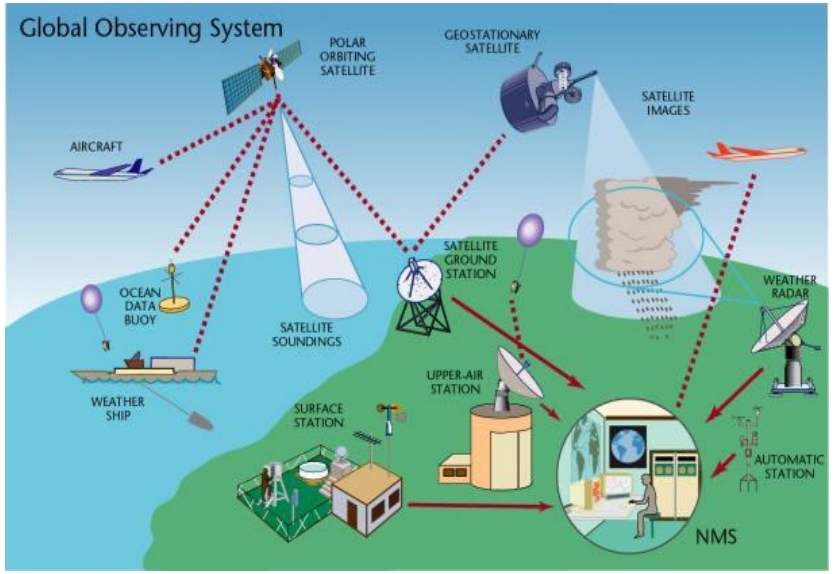
x is a vector of the model variables that we want to estimate.

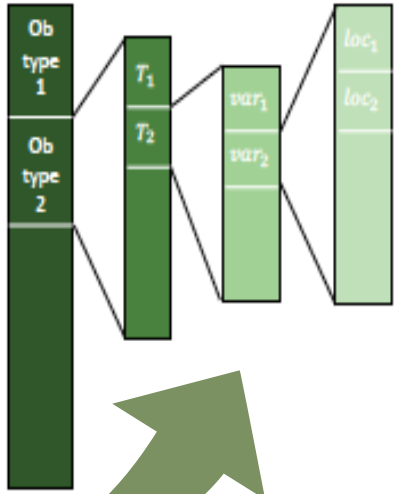
Bayes' theorem

Most DA algorithms can be derived from Bayes' theorem:

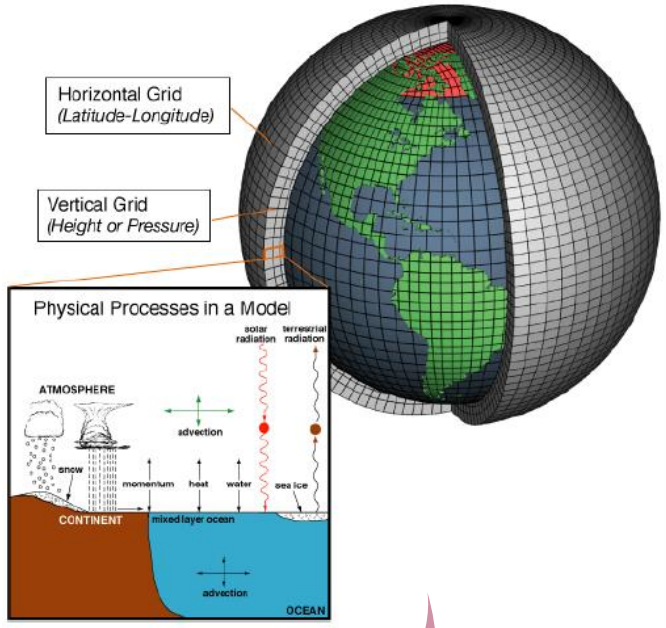
$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$$

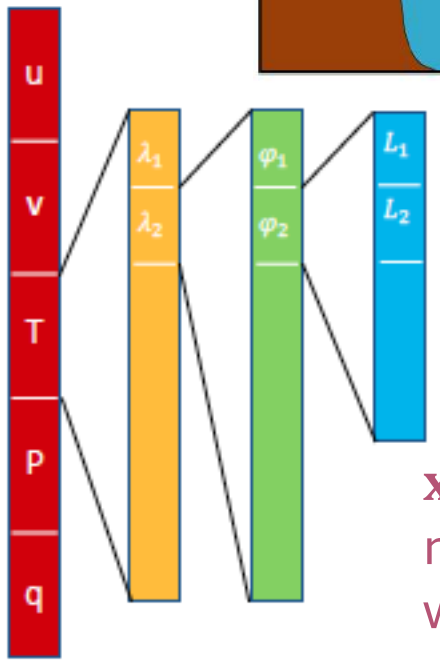
posterior prior likelihood





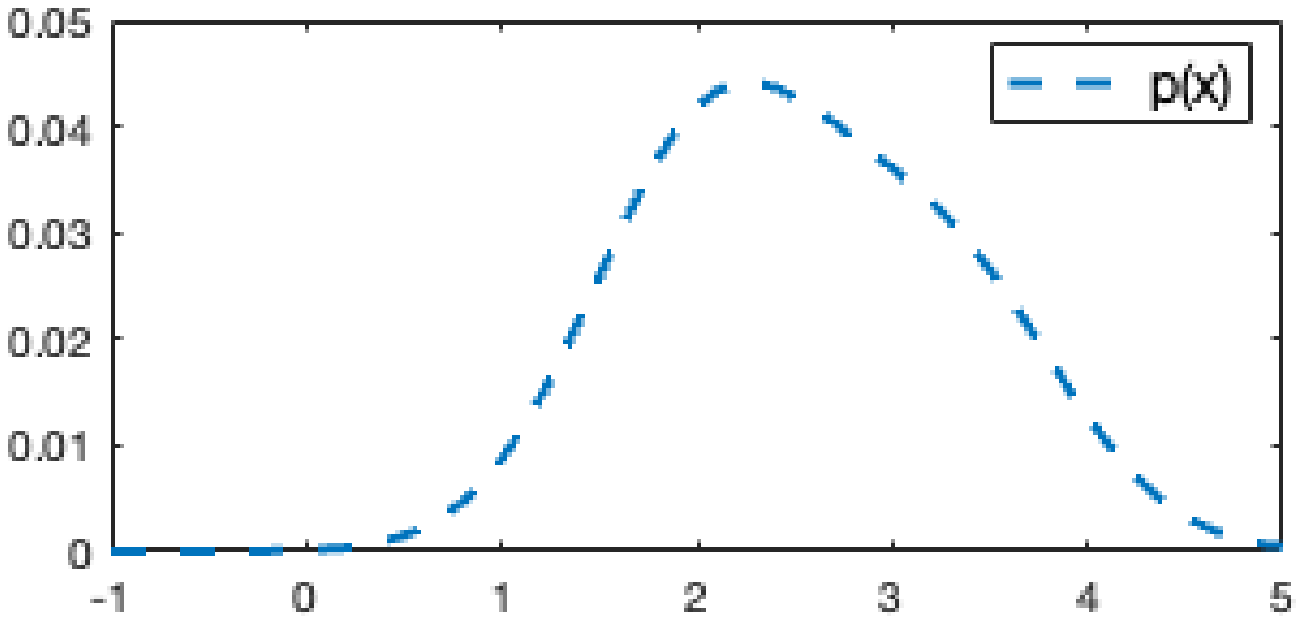
$\mathbf{y}$ is a vector of the available observations





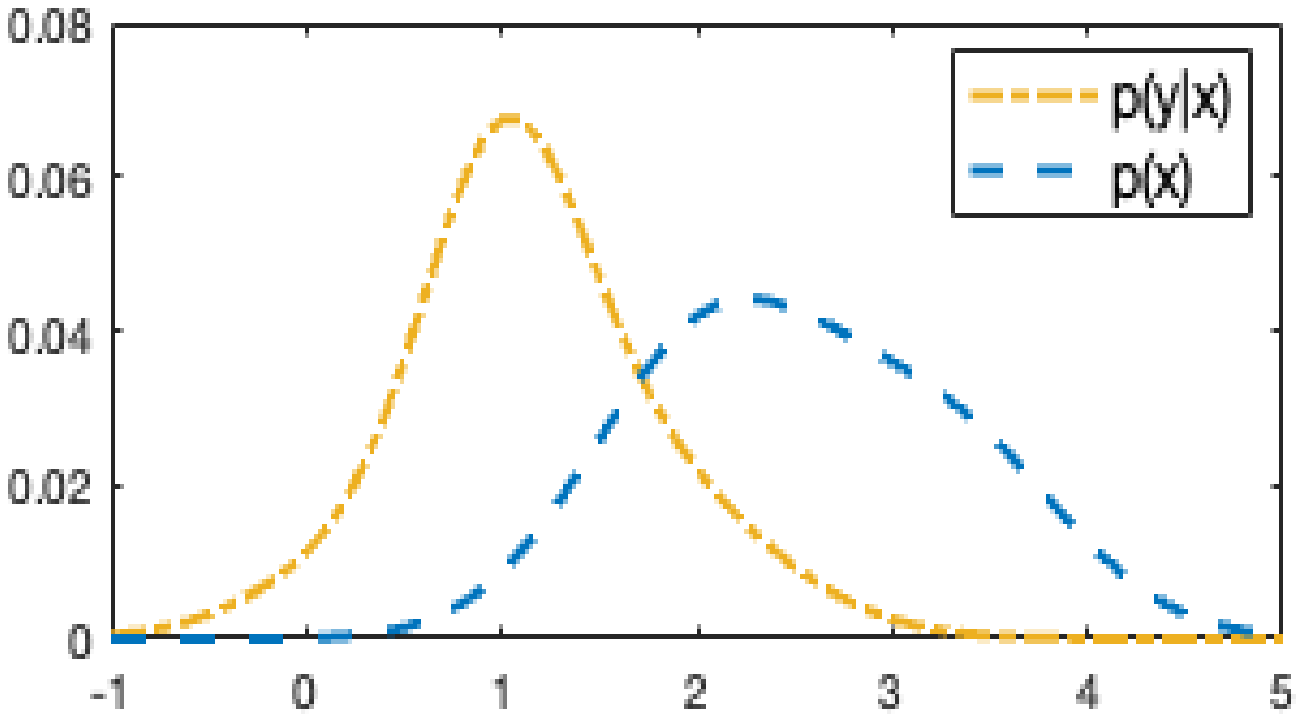
$\mathbf{x}$ is a vector of the model variables that we want to estimate

Bayes' theorem: Scalar illustration $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$



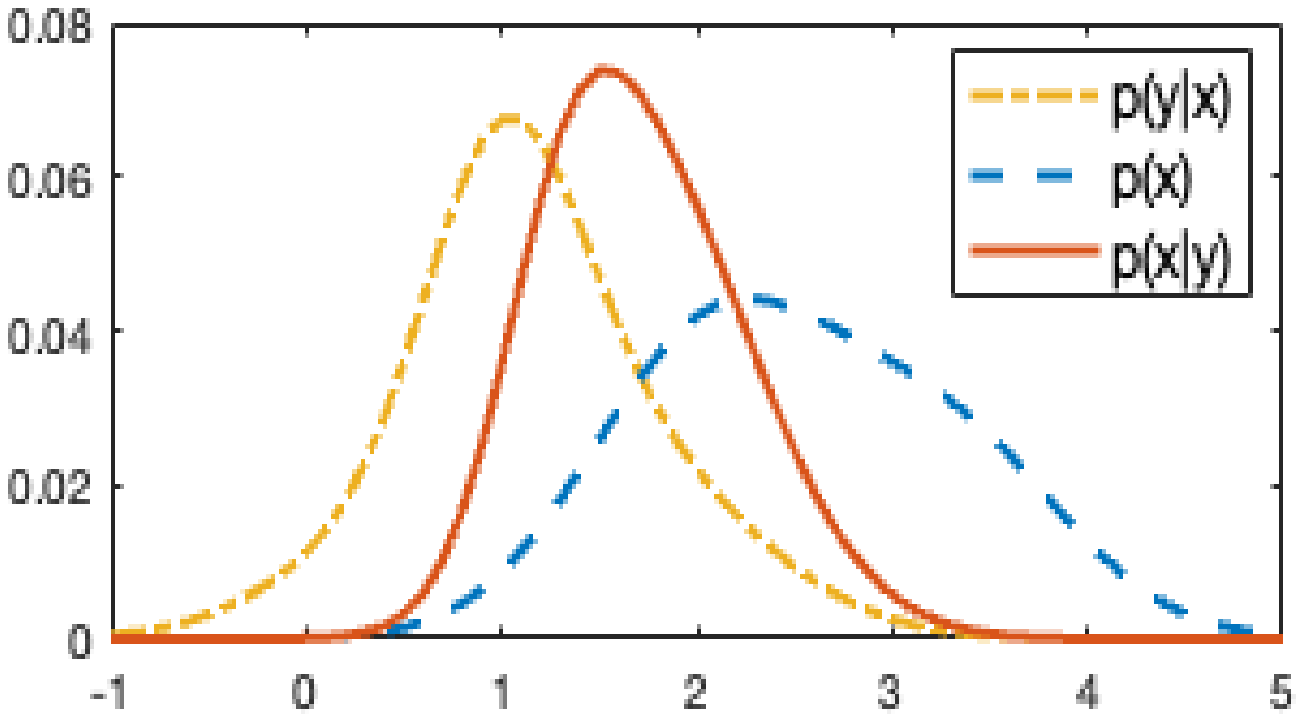
- The prior PDF, p(x), describes the probability of your state variables. Often, this knowledge comes from a previous forecast.

Bayes' theorem: Scalar illustration $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$



- Likelihood PDF, p(y|x), describes the probability of observations given that they are measuring the state we are interested in. P(y|x)=L(x|y) so that we can think of it as a function of x.

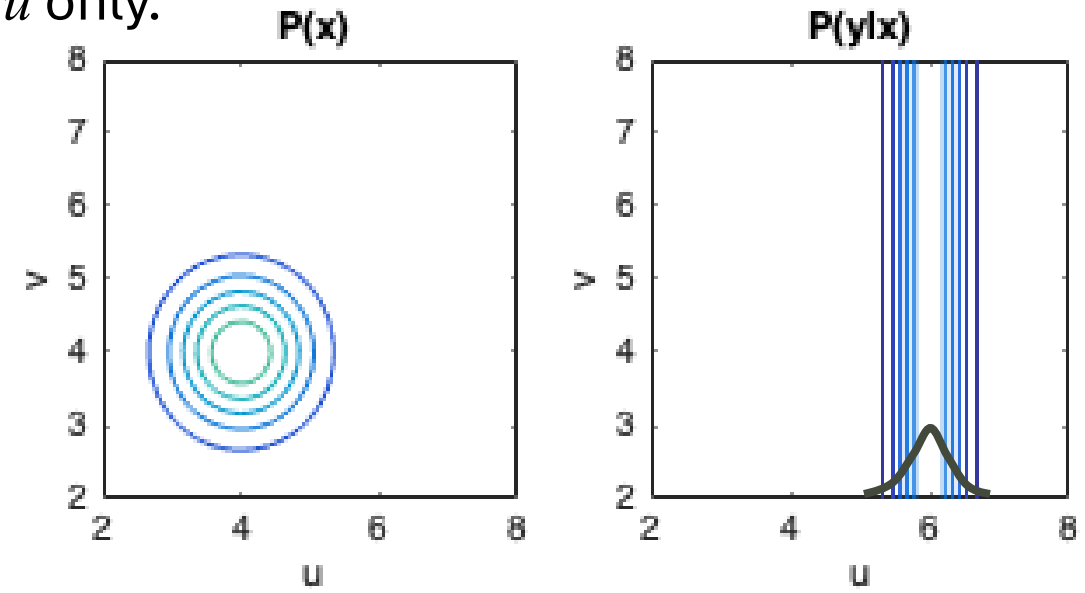
Bayes' theorem: Scalar illustration $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$



- The posterior PDF is given by multiplying the two together and normalising. Updating the prior with information from the observations has shifted the probability mass and reduced the range of probable values of x (i.e. the uncertainty in x is reduced!)

Bayes' theorem: 2 variable example

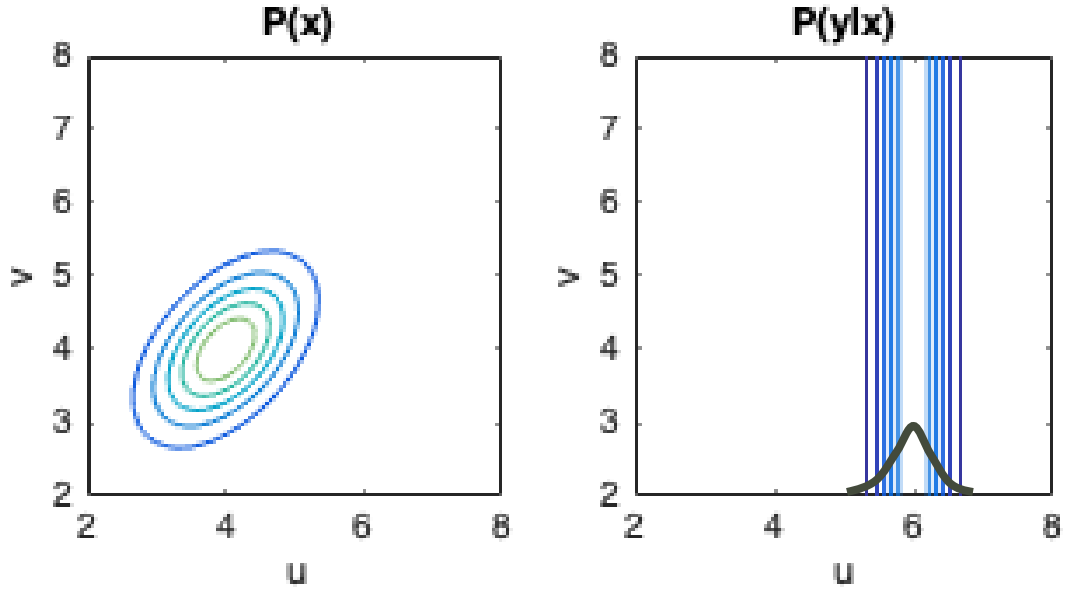
Let our state $\mathbf{x}$ be a vector of zonal and meridional winds, u and v at one location. Observe u only.



Compared to the Prior, the region of high probability in the Posterior for u is reduced and is shifted towards that of the likelihood.

Bayes' theorem: 2 variable example

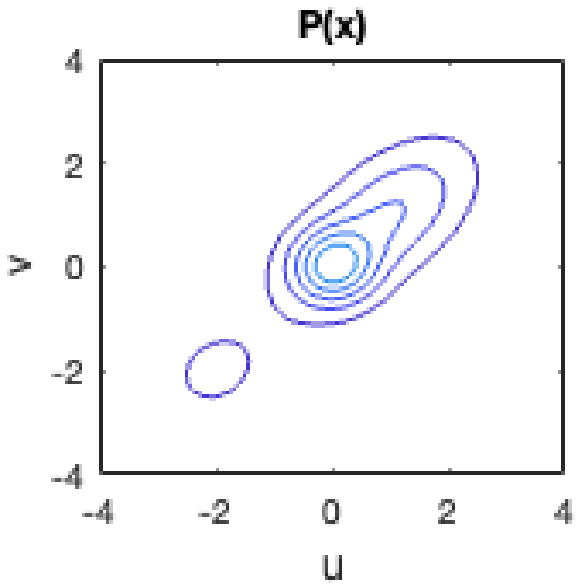
How do you think the Posterior would change if the **Prior** was **correlated**?



When the Prior is correlated, **observations of one variable can be used to update the analysis of both variables**. This helps to **ensure the analysis is physically realistic**.

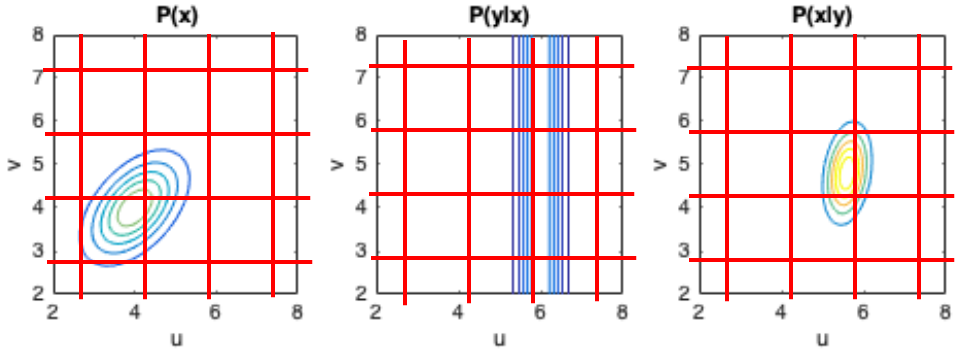
Bayes' theorem: 2 variable example

In this example the prior is non-Gaussian. Let us observe u^2, how will this change the likelihood?



How to solve Bayes' theorem?

A naïve approach could be to discretize the whole of the state space, evaluate $p(x = x_i)$ and $p(y|x = x_i)$ and multiply (this is exactly what I did in the 2D examples)



😕 Might miss the regions of high probability

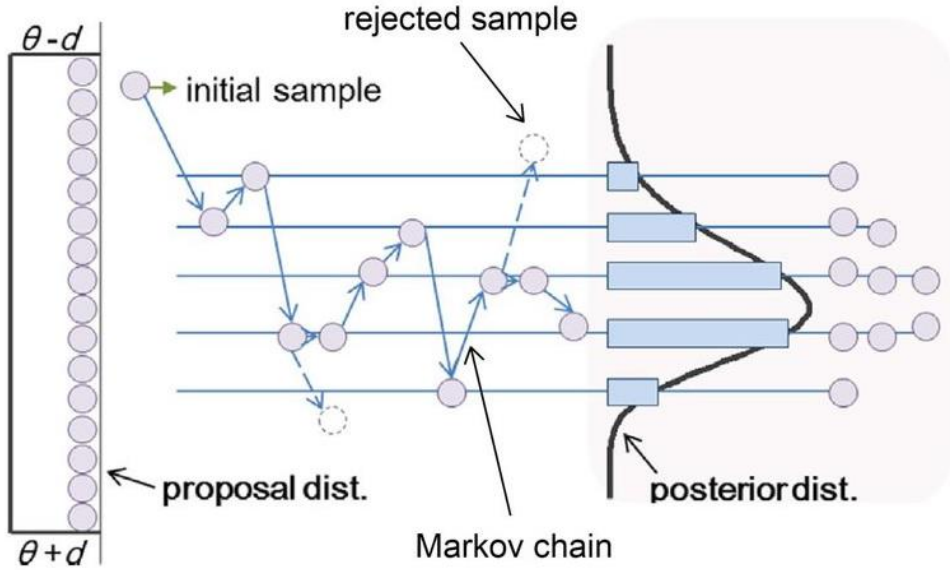
😩 Waste effort evaluating the PDFs in regions of near-zero probability

We will discuss two different, more efficient, approaches to solving Bayes' theorem

1. Non-Parametric approaches: No assumptions are made about the form of Prior or likelihood and consequently the Posterior. These methods include MCMC and the Particle filter
2. Parametric approaches: The prior, likelihood and hence the Posterior are assumed to follow a given distribution. These methods include variational techniques and the EnKF.

Non-parametric approaches

MCMC

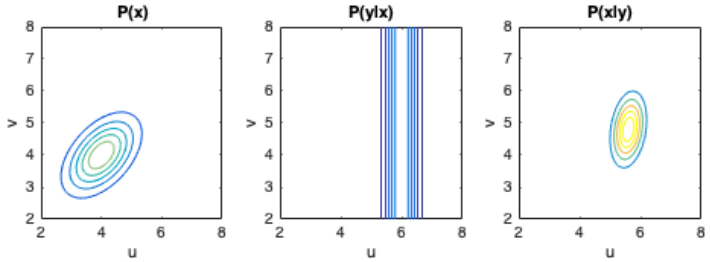


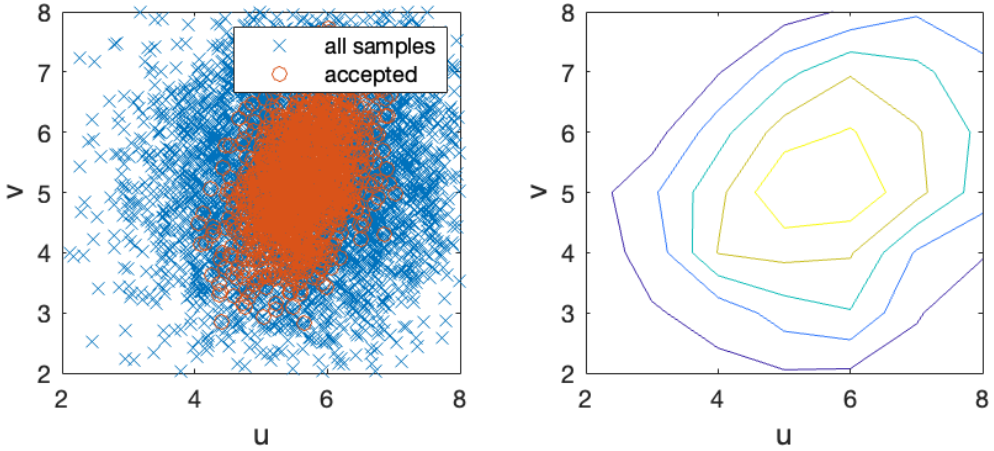
Markov Chain Monte Carlo (MCMC) are a class of algorithms that allow you to sample from the Posterior distribution.

The basic algorithm is:

1. Initialise, $i = 0$, select $\mathbf{x}_i$.
2. Draw a new sample from a proposal distribution $\tilde{\mathbf{x}} \sim q[\tilde{\mathbf{x}}; \mathbf{x}_i]$.
3. Evaluate $\tilde{\mathbf{y}} = h(\tilde{\mathbf{x}})$ and $P(\tilde{\mathbf{x}})$
4. Decide whether to keep $\tilde{\mathbf{x}}$ as part of the sample. Compute $\rho = \frac{P(\tilde{\mathbf{x}})P(\mathbf{y}|\tilde{\mathbf{x}})q[\tilde{\mathbf{x}};\mathbf{x}_i]}{P(\mathbf{x}_i)P(\mathbf{y}|\mathbf{x}_i)q[\mathbf{x}_i;\tilde{\mathbf{x}}]}$ and $\nu \sim U[0,1]$. If $\rho > \nu$ accept proposed sample $\mathbf{x}_{i+1} = \tilde{\mathbf{x}}$, else it is rejected.
5. Set i=i+1 and return to 2.

MCMC



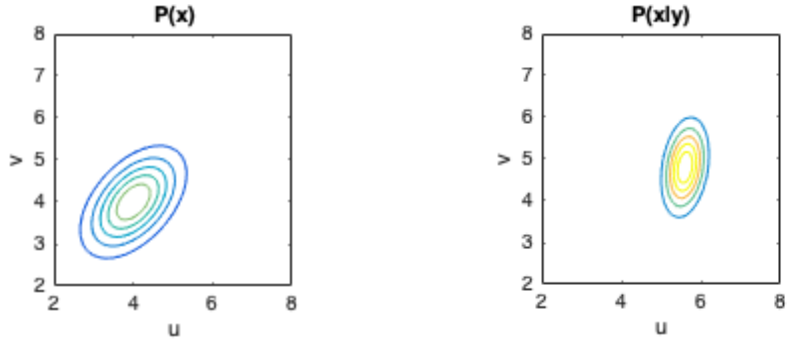


The accept/reject procedure ensures that samples that provide a better fit to the observations are immediately accepted, those that provide a similar fit are considered, and those that lead to simulated observations that are very different from the measurements are rejected.

Regions with relatively high probability are therefore preferentially sampled, whereas regions with low probability are avoided, and a sample of the posterior distribution is produced using far fewer iterations than direct computation of the PDF.

The theoretical underpinnings of the MCMC algorithm can be found in Mosegaard and Tarantola(2002) and Tarantola (2005).

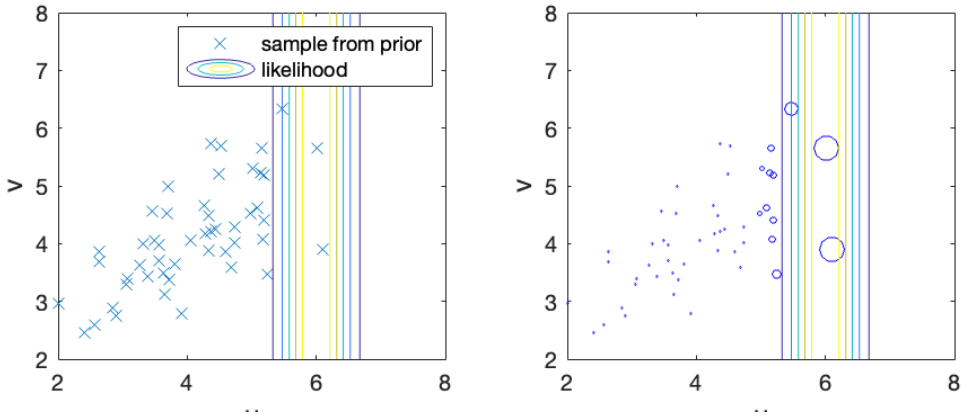
The Particle filter



Like the MCMC the Particle filter (PF, also known as sequential MC) aims to produce a sample representation of the posterior distribution.

The basic idea is:

1. Draw a sample from the prior distribution $\mathbf{x}_i \sim P(\mathbf{x})$, for $i = 1, ...N$.
2. For each sample evaluate $\mathbf{y}_i = h(\mathbf{x}_i)$.
3. For each sample compute $w_i = P(\mathbf{y}|\mathbf{x}_i)$.
4. Normalise the weights $w_i = w_i / \sum_i w_i$



Prior sample

Posterior sample

The posterior is then given by

$$P(\mathbf{x}|\mathbf{y}) \approx \sum_i w_i \delta(\mathbf{x} = \mathbf{x}_i)$$

Application of Bayes' to large problems

Non-parametric methods that make no assumption about the nature of the prior and the likelihood are inefficient for large-scale problems because the number of samples needed to represent the whole distribution grows exponentially with the size of the state to be estimated.

CURSE OF DIMENSIONALITY

The efficiency of the MCMC method can be quantified by the acceptance/reject ratio (0.33 in the example).

The efficiency of the particle filter can be measured by the effective sample size:

$$ess = \frac{1}{\sum_{i=1}^{N} w_i^2}$$

If the weights are very uneven the ess can approach 1 and the weighted sample will be overwhelmed by sampling noise (4.7 with N=50 in the example).

To increase the ess, PFs are being developed to increase the chance that the sample is the region of high likelihood, see van Leeuwen et al. 2019.

Parametric approaches - Gaussian assumption

In NWP we are interested in applying Bayes' theorem to approximately 10^8 dimensions.

In many cases, it is appropriate to assume that the prior and likelihood are Gaussian

$$f(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Mean vector

Error covariance matrix

Gaussian assumption

The prior:

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

Mean of the prior distribution

Prior error covariance matrix

The likelihood:

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

Vector of observations

observation error covariance matrix

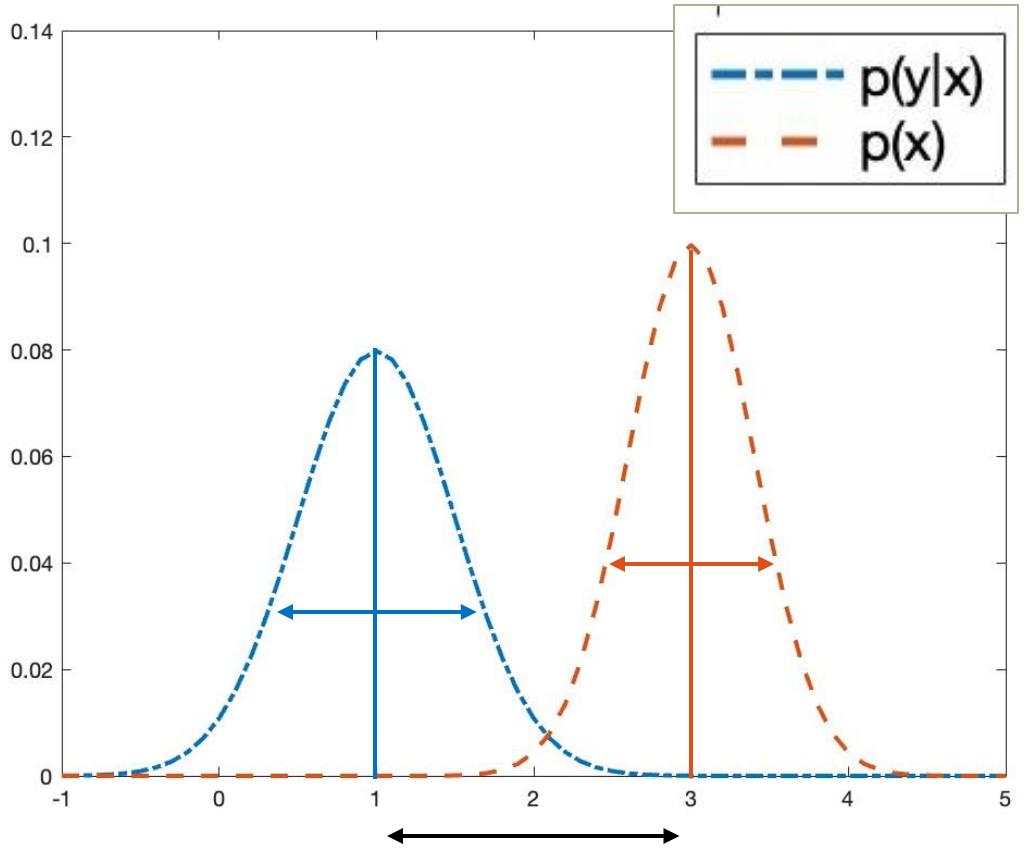
Observation operator, mapping from state to observation space

Applying Bayes' theorem the posterior is:

$$p(\mathbf{x}|\mathbf{y}) \propto exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}^{\mathrm{b}})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^{\mathrm{b}}) - \frac{1}{2}(\mathbf{y}-h(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-h(\mathbf{x}))\right)$$

If *h* is linear then the posterior is also Gaussian and can be parameterized according to its mean (xa, the analysis) and its (analysis error) covariance matrix.

Bayes' theorem scalar Gaussian illustration



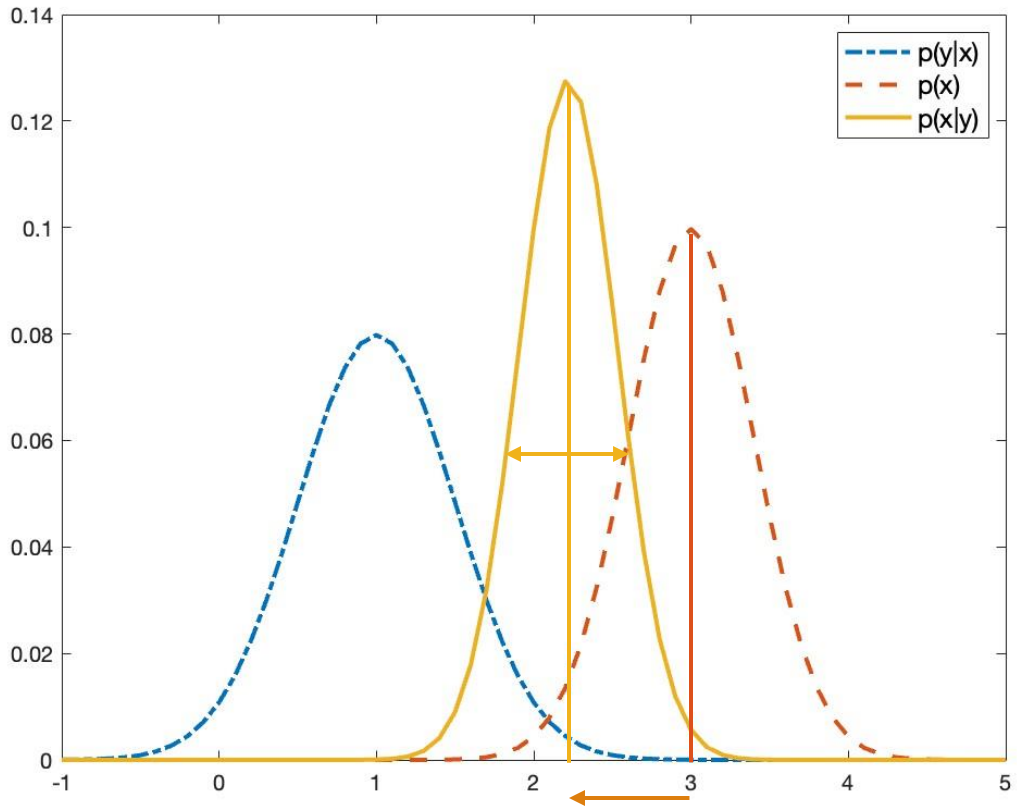
Innovation $= y - h(x^b)$

e.g. temperature °C

The **mean** of p(y|x) and p(x) are the **observed** and **background/forecast** model values respectively.

The **standard deviations** of p(y|x) and p(x) are the **uncertainties** of the observed and background values respectively.

Bayes' theorem scalar Gaussian illustration



Analysis increment

The uncertainty of the posterior is smaller than either the likelihood or the prior.

The mean = mode of p(x|y) gives the '**analysis**', which is also the minimum variance estimate!

Analytical solution – the Kalman equations

The analysis (the maximum a posteriori state) can be derived analytically as

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{K}(\mathbf{y} - h(\mathbf{x}^{\mathrm{b}})) \qquad \text{where} \qquad \mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

and $\mathbf{H}$ is the linearized observation operator.

$\mathbf{K}$ (known as the Kalman gain) prescribes the weight given to the observations versus the prior. We see that the $\mathbf{K}$ increases as the prior uncertainty ($\mathbf{B}$) increases and the observation uncertainty ($\mathbf{R}$) decreases.

The analysis error covariance can also be derived analytically

$$\mathbf{P}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}$$

We see as $\mathbf{K}$ increases $\mathbf{P}^{\mathrm{a}}$ decreases.

In practice, we can not evaluate these expressions directly.

Variational and Ensemble Kalman Filter techniques

Variational (Var) DA

Var algorithms aim to find Maximum a-posterioiri (MAP) state/parameter set, which is also the mean of the Posterior distribution (assuming Gaussianity) and hence the same as the minimum variance estimate (Lorenc 1986).

In Var the analysis (the MAP estimate) is found by minimising the following cost function:

$$P(\mathbf{x}|\mathbf{y}) \sim N(\mathbf{x}^a, \mathbf{P}^a)$$

$$P(\mathbf{x}) \sim N(\mathbf{x}^b, \mathbf{B})$$

$$P(\mathbf{y} \mid \mathbf{x}) \sim N(\mathbf{y}, \mathbf{R})$$

$$\begin{aligned} J(\mathbf{x}) &= -const. \ln P(\mathbf{x}|\mathbf{y}) \\ &= -\ln(P(\mathbf{x})P(\mathbf{y}|\mathbf{x})) \\ &= -\ln P(\mathbf{x}) - \ln P(\mathbf{y}|\mathbf{x}) \end{aligned}$$

$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + (\mathbf{y}-h(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-h(\mathbf{x}))$$

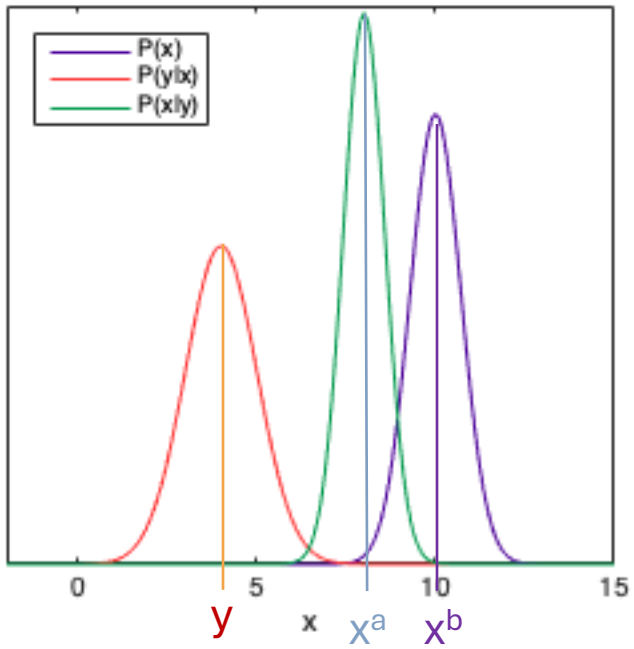
J_b + J_o

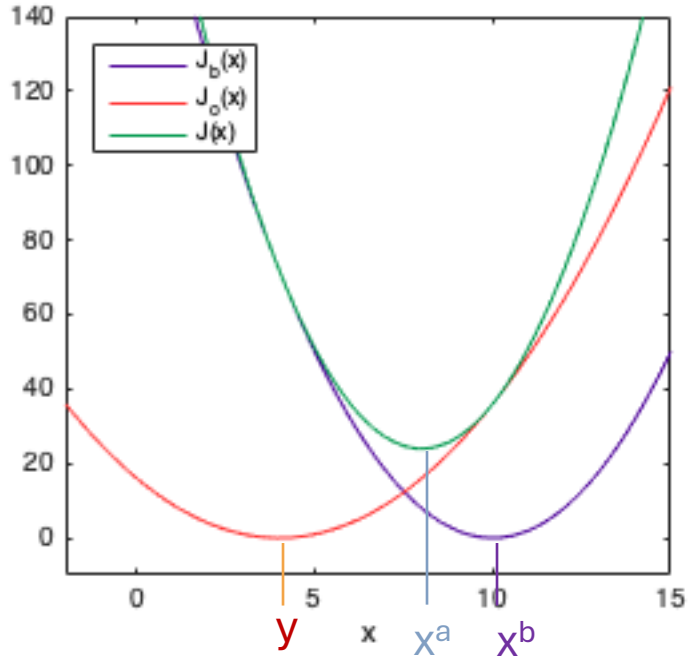
If the observation operator is linear then the cost function is quadratic

Variational DA

1 variable

Example:

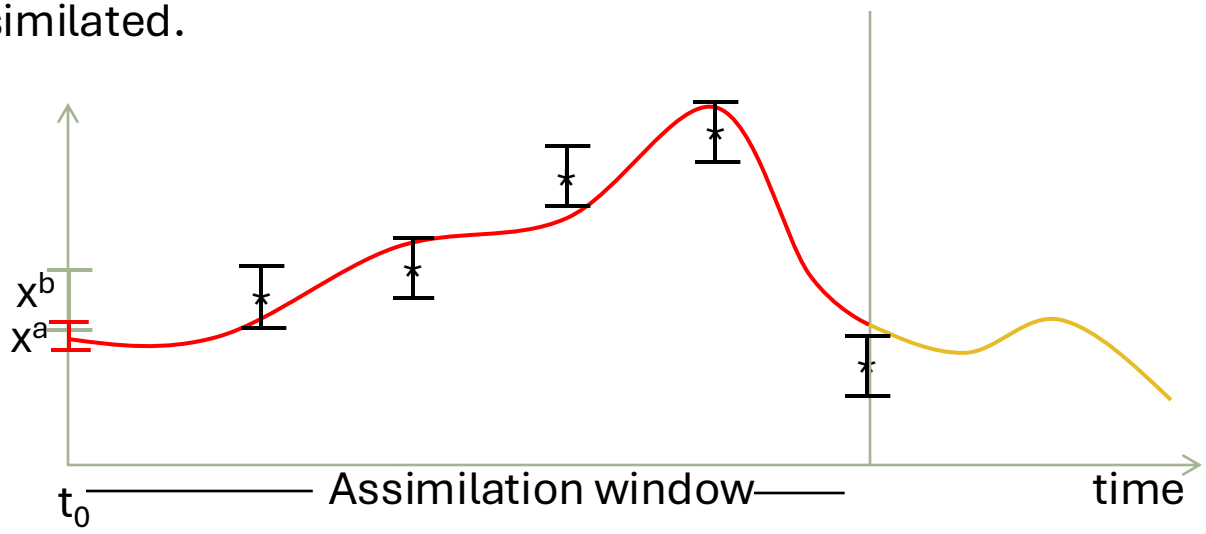




The minimum of the cost function, $\mathbf{x}^a$, is known as the analysis.

The assimilation window (4DVar)

The general form of variational DA is called 4DVar. This allows for observations over a window to be assimilated.



*Observations

background uncertainty, characterised by **B**

observation uncertainty, characterised by **R**

$M(x^a)$

analysis uncertainty.

forecast

The 4DVar cost function is:

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{i=1}^{N}(\mathbf{y}_i - h(m(\mathbf{x}_0, t_0, t_i))^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{y}_i - h(m(\mathbf{x}_0, t_0, t_i))$$

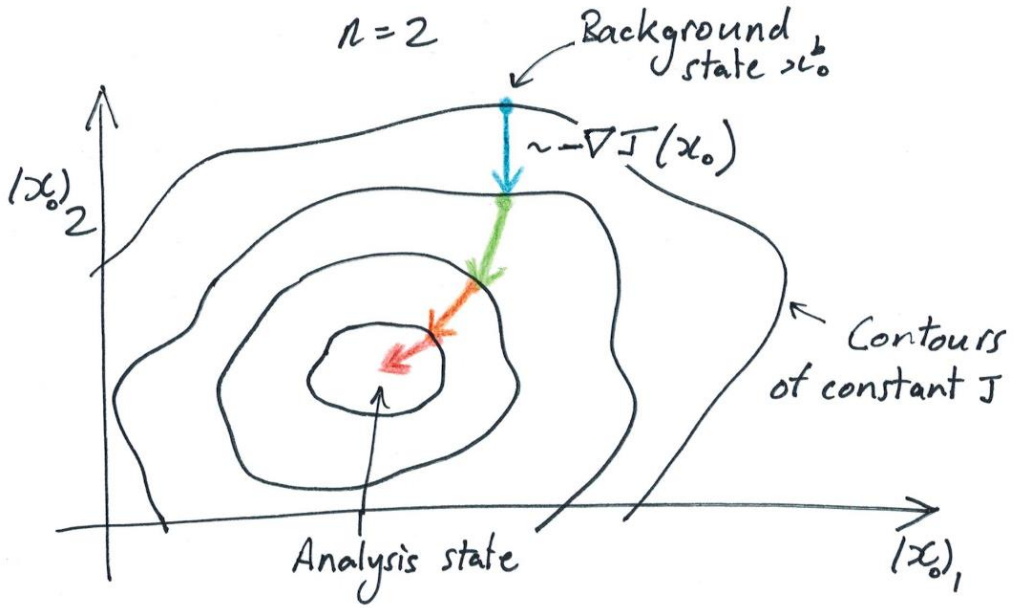
Can think of $h(m(\mathbf{x}_0, t_0, t_i)$ as a generalized ob operator:

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + (\hat{\mathbf{y}} - \hat{h}(\mathbf{x}_0))^{\mathrm{T}}\hat{\mathbf{R}}^{-1}(\hat{\mathbf{y}} - \hat{h}(\mathbf{x}_0))$$

4DVar

The minimum of the cost function (the analysis) can be found iteratively by searching in the direction of the gradient

$$\mathrm{x}_0^{k+1} = \mathrm{x}_0^k + \alpha \nabla J(\mathrm{x}_0^k)$$



The gradient of the cost function is given by

$$\nabla \mathrm{J}\left(x_0^k\right) = \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^k\right) + \sum_{i=1}^{p} \mathbf{M}_{t_0 \to t_i}^{\mathrm{T}} \quad \mathbf{H}^{\mathrm{T}} \mathbf{R}_i^{-1}\left(\mathbf{y}_i - h(M_{t_0 \to t_i}(\mathbf{x}_0)\right),$$

where $\mathbf{M}$ is the tangent linear of the forecast model, $\mathbf{M} \in \mathbb{R}^{n \times n} = \frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_0}$ evaluated at $\mathbf{x}_0^k$. $\mathbf{M}^{\mathrm{T}}$ is the model adjoint.

An alternative approach

A major disadvantage of variational techniques is that $\mathbf{B}$ is generally not updated on each assimilation cycle. In variational DA $\mathbf{B}$ is only designed to represent a climatological estimate of the error covariances. However, the errors can be highly flow-dependent.

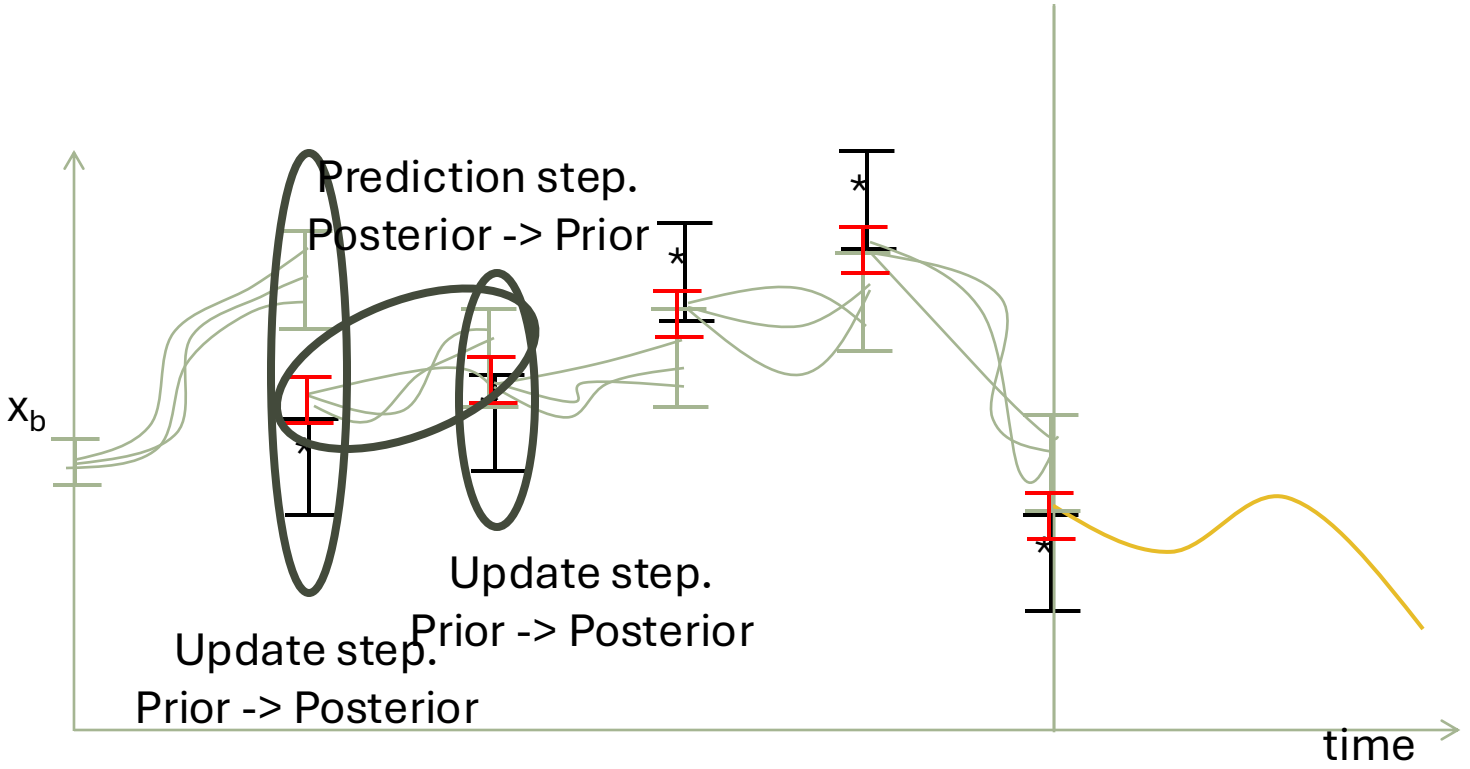
This has motivated the development of the Ensemble Kalman Filter (EnKF).

The EnKF also based on linear/Gaussian theory so is a retractable method for large systems.

Unlike Var the EnKF represents $\mathbf{B}$ using an ensemble (and calls it $\mathbf{P}^{\mathrm{f}}$).

The ensemble Kalman filter (EnKF)

The EnKF (Evensen 1994) merges KF theory with Monte Carlo estimation methods



The ensemble Kalman filter (EnKF)

The EnKF makes use of an ensemble approximation of the forecast error covariance matrix (previously the B matrix) to allow us to compute the Kalman equations directly.

$$\mathbf{P}^f = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_k^{(i),\mathrm{f}} - \bar{\mathbf{x}}_k^{\mathrm{f}}\right)\left(\mathbf{x}_k^{(i),\mathrm{f}} - \bar{\mathbf{x}}_k^{\mathrm{f}}\right)^{\mathrm{T}} = \frac{1}{N-1}\mathbf{X}'^f(\mathbf{X}'^f)^T$$

Number of ensemble members

Forecast perturbation matrix . Dimension $n \times N$

The update step for the mean: $\bar{\mathbf{x}}^{\mathrm{a}} = \bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{K}(\mathbf{y} - h(\bar{\mathbf{x}}^{\mathrm{f}}))$, $\mathbf{K} = \mathbf{P}^f\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\boldsymbol{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$

EnKF – the update step for the covariance

The analysis error covariance matrix can be estimated in two ways

- **Stochastic methods**: Samples from the posterior distribution are created

$$\mathbf{x}_k^{(i),\mathrm{a}} = \mathbf{x}_k^{(i),\mathrm{f}} + \mathbf{K}_k(\boldsymbol{y}_k + \boldsymbol{\epsilon}_{\mathrm{y}}^{(i)} - h\left(\mathbf{x}_k^{(i),\mathrm{f}}\right)), \text{ where } \boldsymbol{\epsilon}_{\mathrm{y}} \sim N(\mathbf{0}, \mathbf{R}).$$ Then the sample covariance matrix is computed as:

$$\mathbf{P}^a = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_k^{(i),a} - \bar{\mathbf{x}}_k^{\mathrm{a}}\right)\left(\mathbf{x}_k^{(i),a} - \bar{\mathbf{x}}_k^{\mathrm{a}}\right)^{\mathrm{T}}$$

- **Deterministic methods**: An analysis perturbation matrix is generated by applying a transformation to the forecast perturbation matrix $\mathbf{X}_k'^{\mathrm{a}} = \mathbf{X}_k'^{\mathrm{f}}\mathbf{T}_k$. The transform $\mathbf{T}$ is derived to ensure $\mathbf{P}^{\mathrm{a}} = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f$ holds.

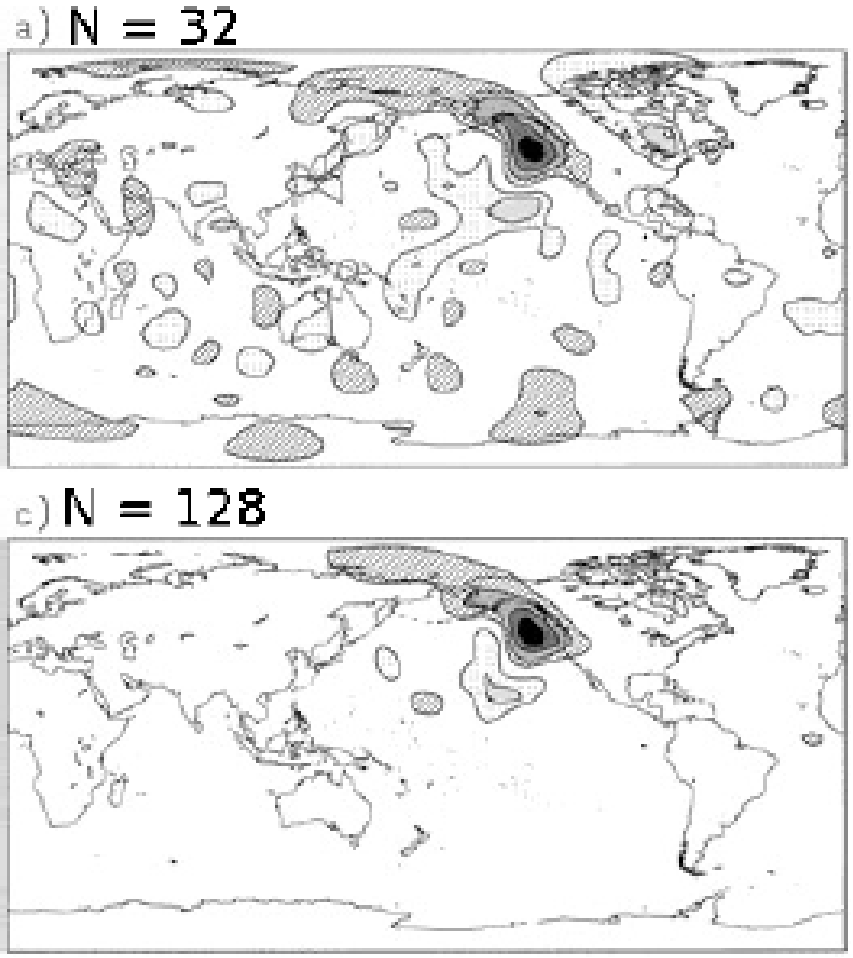
$$\mathbf{P}^a = \frac{1}{N-1}\mathbf{X}'^a(\mathbf{X}'^a)^T$$

EnKF and Sample errors

Because the ensemble size is limited, $N \ll n$. This means that the ensemble estimate of $\mathbf{P}^{\mathrm{f}}$ will be low rank and affected by sample errors.

The sample errors will affect

- The estimate of the **error variances** so that the wrong weight is given to the observations. When the the error variances are underestimated this can lead to filter divergence!!!!
- The **error correlations** such that observations can update variables which they do not contain information about.



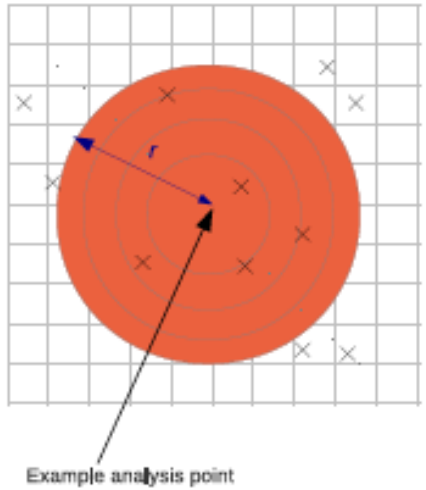
From Houtekamer & Mitchell (1998)

EnKF and Sample errors

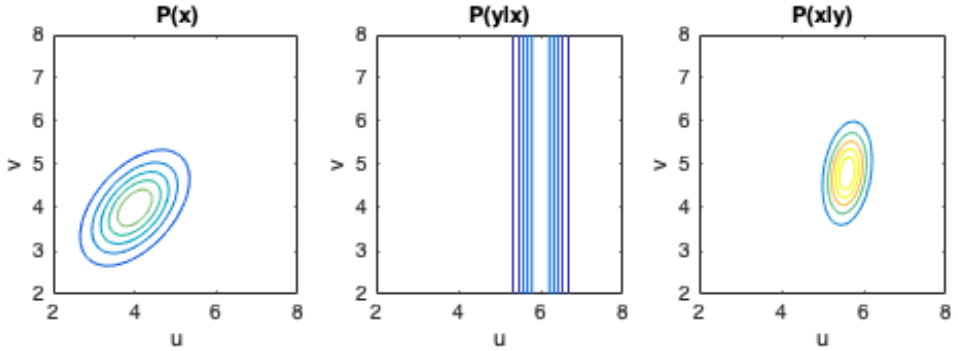
To mitigate these two problems, we use a combination of

- **variance inflation** (e.g. Mitchell and Houtekamer, 2000; Anderson and Anderson, 1999, Whitaker and Hamill, 2012) and

- **covariance localization (**Houtekamer and Mitchell, 2001, Hunt et al. 2007).



Summary

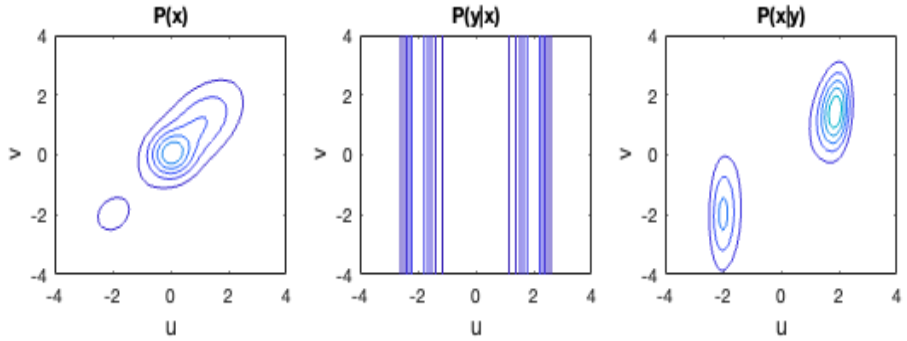


There are many different data assimilation algorithms proposed. The ones presented all aim to solve Bayes' theorem to find the probability distribution of the state consistent with the uncertainty in the observations and model.

Many methods used operationally rely on the Gaussian assumption to make the DA problem tractable for large dimensions and efficient to run as part of a cycled forecasting system.

These methods include the variational and Ensemble Kalman Filter techniques.

Summary



However, there are many reasons why the Gaussian assumption may not hold:

- Non-linear model
- Non-linear observation operator
- Non-Gaussian errors e.g. if the variable is bounded

Assuming Gaussianity in these cases can result in:

- High probabilities assigned to unphysical states that may result in numerical instabilities
- Low probabilities assigned to important regimes
- Extreme events not represented

Even within Var and EnKF, there is room for the relaxation of the Gaussian assumption e.g. inner/outer loops, Gaussian anamorphosis.

Can reformulate the problem to not make Gaussian assumptions e.g. the particle filter.

A note on parameter estimation

Each of the DA methods can be modified to estimate parameters instead of (or as well as) estimating the initial state. For example,

- Model parameters describing physical processes
- Parameters to describe the bias correction of the model or observations

Need to consider

- How the prior uncertainty is represented. If you updating the parameters using variational or EnKF methods, does a Gaussian error make sense?
- How can the observations be related to these parameters? Often, you do not observe parameters directly, so instead rely on their errors being correlated with state variables that are observed.

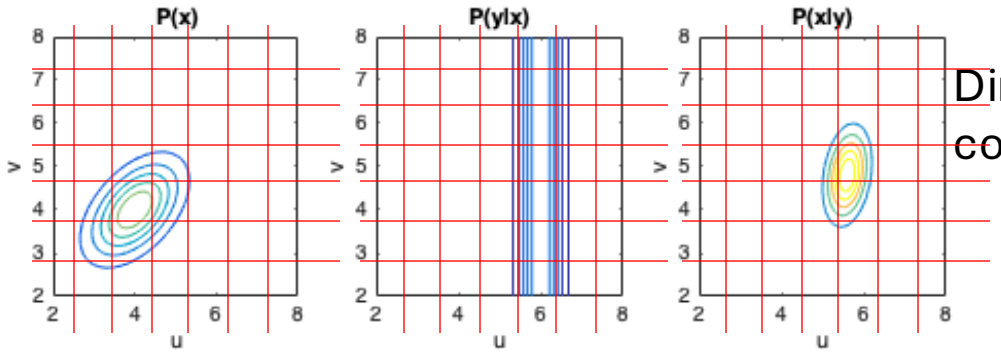
DA software

On Wednesday and Thursday we will dedicate more time to learning about Variational and EnKF algorithms with computer practicals.

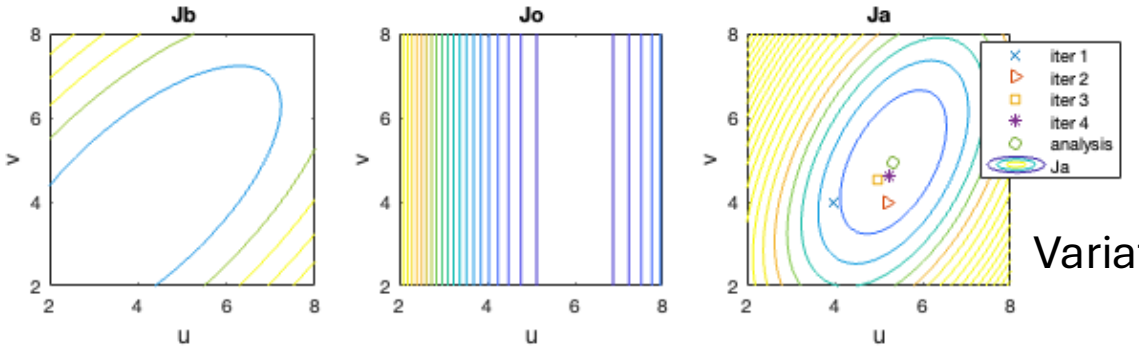
Our code is available on https://github.com/darc-reading/darc-training-2025/

Tomorrow Yumeng and Chris will give a lecture DA software packages

Schematic comparison of techniques

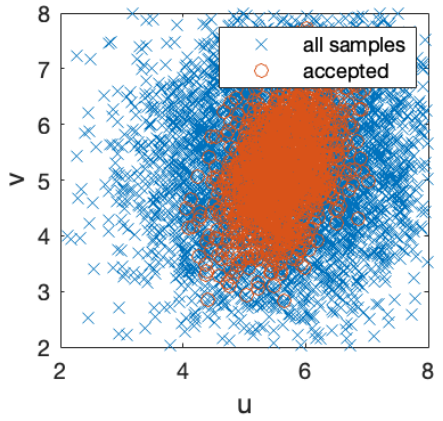


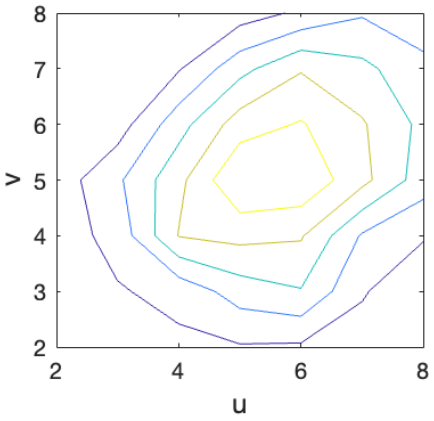
Direct computation



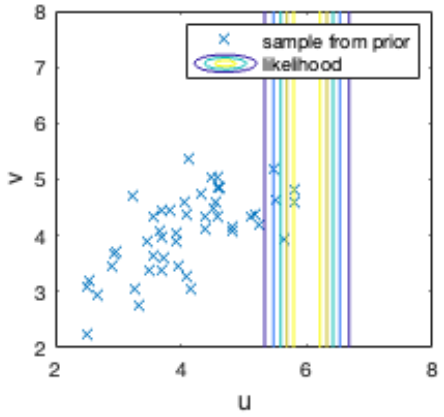
Variational

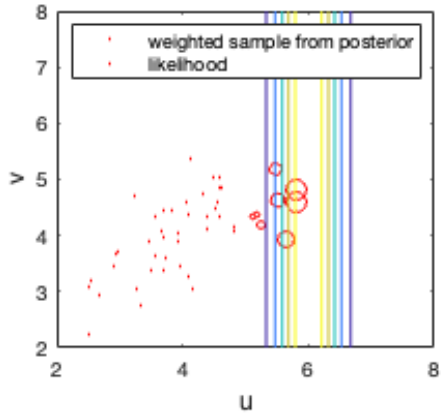
Can combine the best bits of (hybridise) the different algorithms



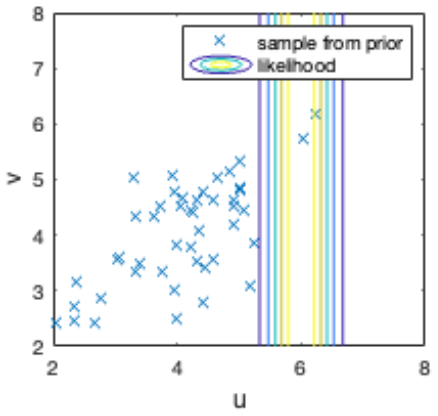


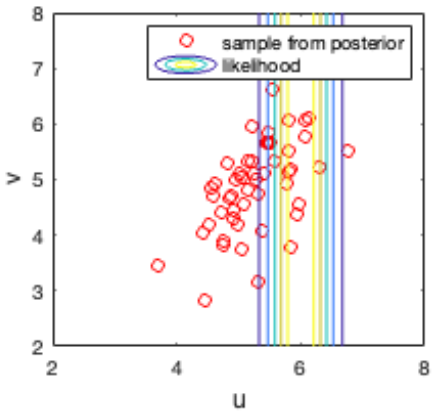
MCMC





Particle Filter





EnKF

References

- Mosegaard and Tarantola (2002): Probabilistic Approach to Inverse Problems. International Geophysics
- Tarantola (2005): *Inverse problem theory and methods for model parameter estimation*. SIAM
- van Leeuwen et al. (2019): Particle filters for high-dimensional geoscience applications: A review. *Q J R Meteorol Soc*.
- Lorenc (1986): Analysis methods for numerical weather prediction. Q.J.R. Meteorol. Soc
- Evensen (1994): Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*
- Mitchell and Houtekamer, (2000): An Adaptive Ensemble Kalman Filter. *Mon. Wea. Rev.*
- Anderson and Anderson, (1999): A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Mon. Wea. Rev.*
- Whitaker and Hamill, (2012) Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Mon. Wea. Rev.*
- Houtekamer and Mitchell, (2001): A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*
- Hunt et al. (2007): Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D: Nonlinear Phenomena