Introduction to Data Assimilation for paleo climate

Peter Jan van Leeuwen
Data Assimilation Research Center (DARC)
National Centre for Earth Observation
Department of Meteorology
University of Reading
p.j.vanleeuwen@reading.ac.uk

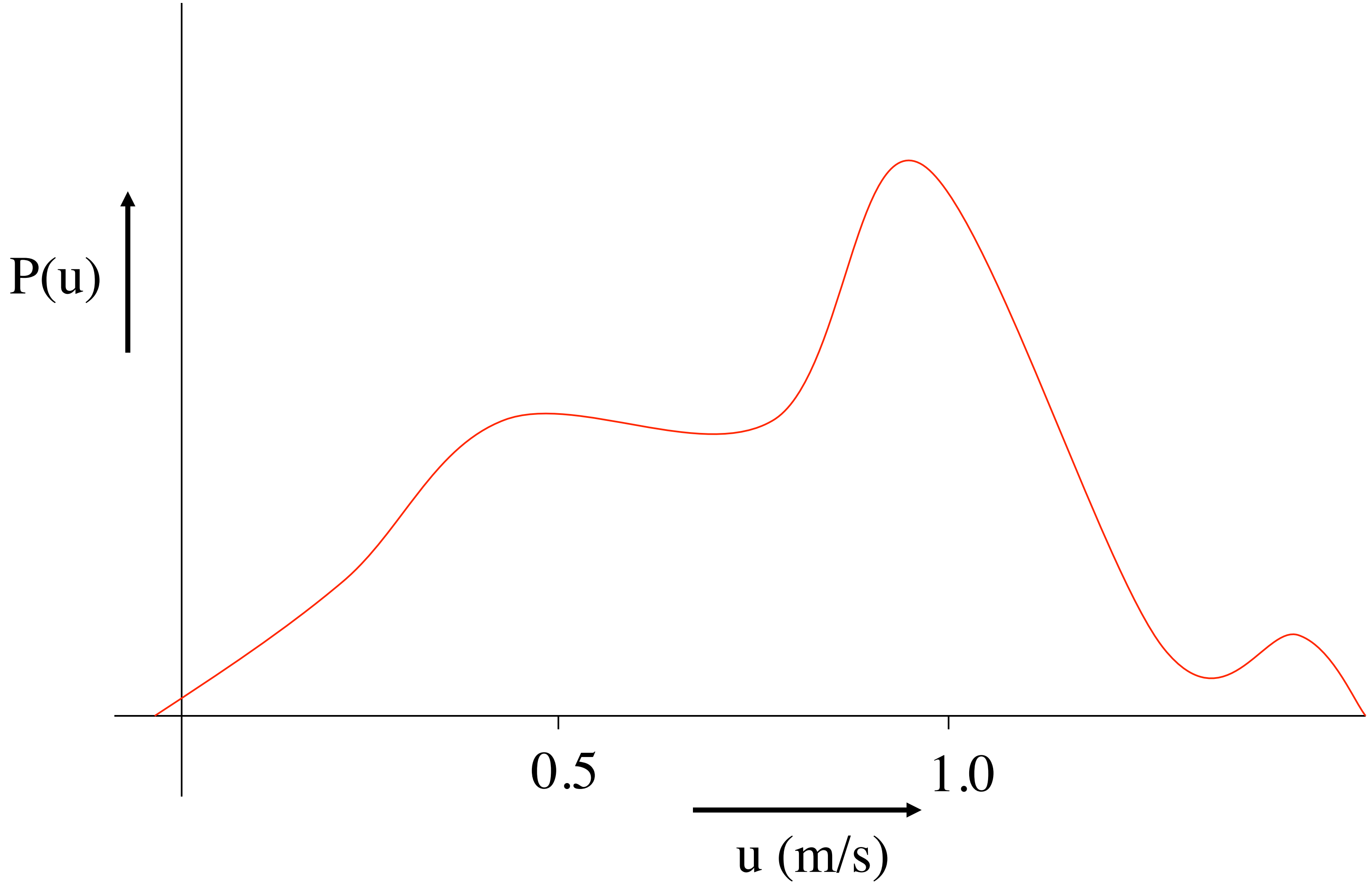
What is data assimilation?

- We have information from the system via physical (chemical,…) knowledge encoded in our numerical model and perhaps other constraints, the *prior*,
- We have information from the system via observations, the *likelihood*,
- Data assimilation is the mathematics of combining the two, in the *posterior*.

Why data assimilation?

- Forecasts
- Process studies via ‘reanalyses’
- Model improvements
  - model parameters
  - parameterizations
- ‘Intelligent monitoring’

The basics: probability density functions

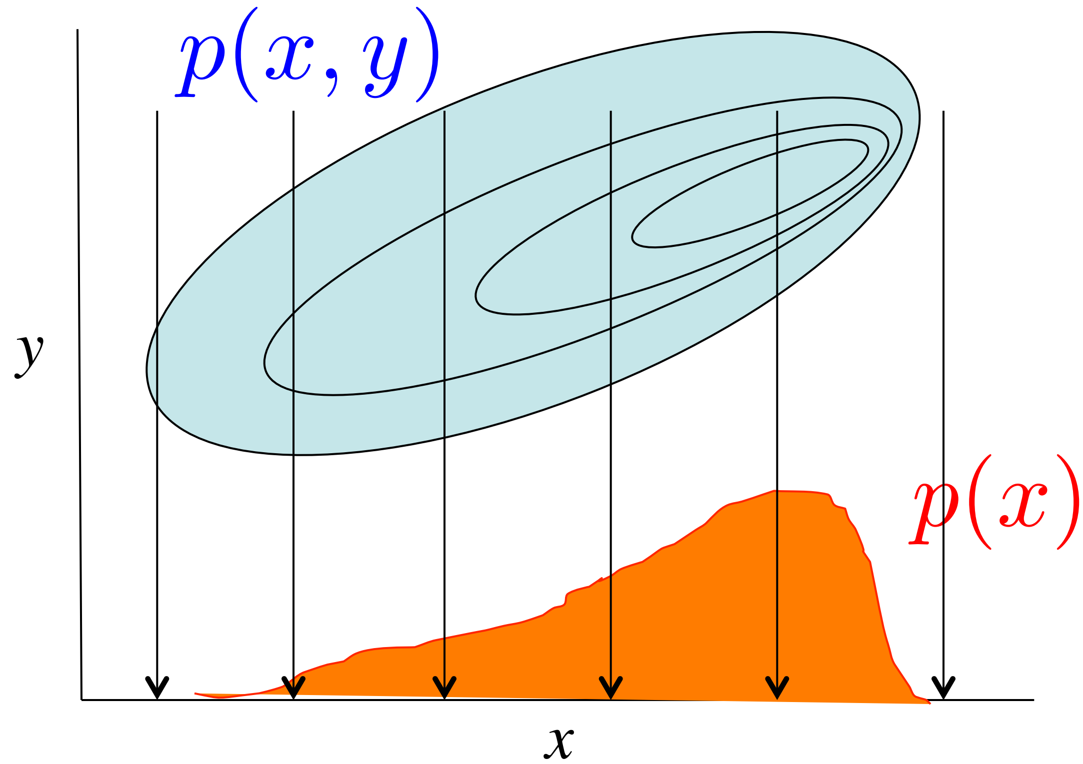


Joint and marginal pdf

Any pdf of more than one variable is called a *joint pdf.*
If we are only interested in the pdf of one of these variables, or of a subset of these variables we can form the *marginal pdf* by integrating the joint pdf over all variables not of interest:

$$p(x) = \int p(x, y)\, dy$$

Integration over y



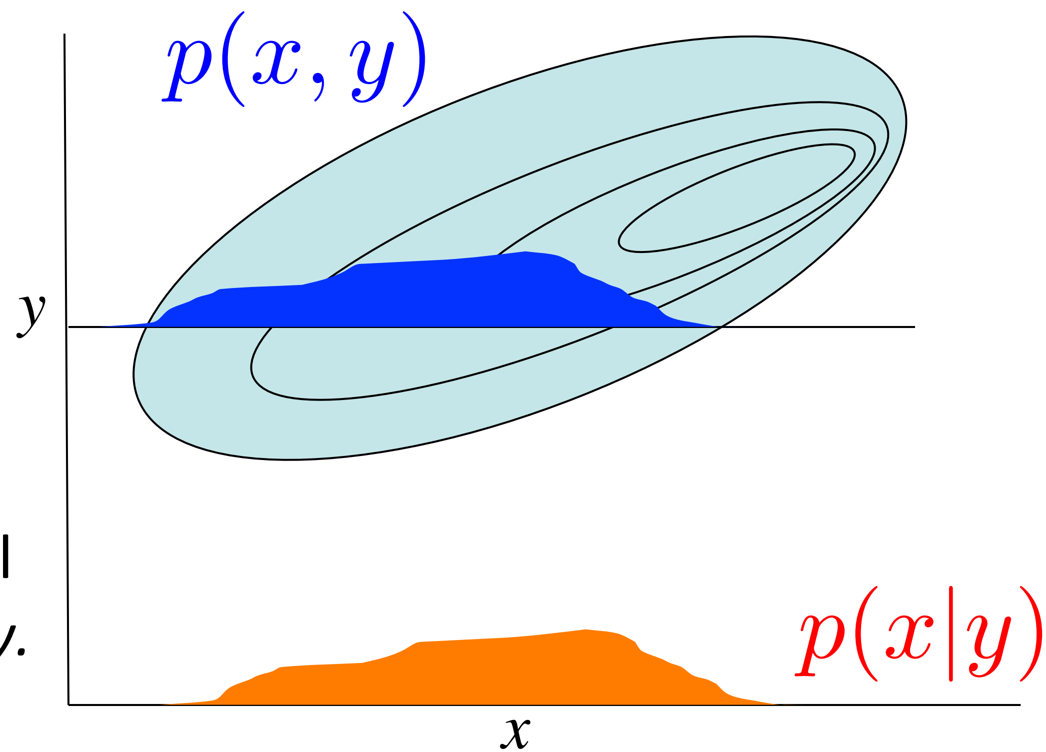
Conditional pdf

We can also form a conditional pdf from a joint pdf by keeping one, or a subset, of the variables constant: $p(x|y)$.

We see that this pdf is equal to $p(x,y)$ along the line *y=y*. However, it is not normalised to 1, unless we define

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

Indeed, *p(y)* is the integral of *p(x,y)* along the line *y=y*.



Bayes Theorem

Conditional pdf:

$$p(x, y) = p(x|y)p(y)$$

Similarly:

$$p(x, y) = p(y|x)p(x)$$

Combine:

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

Bayes Theorem

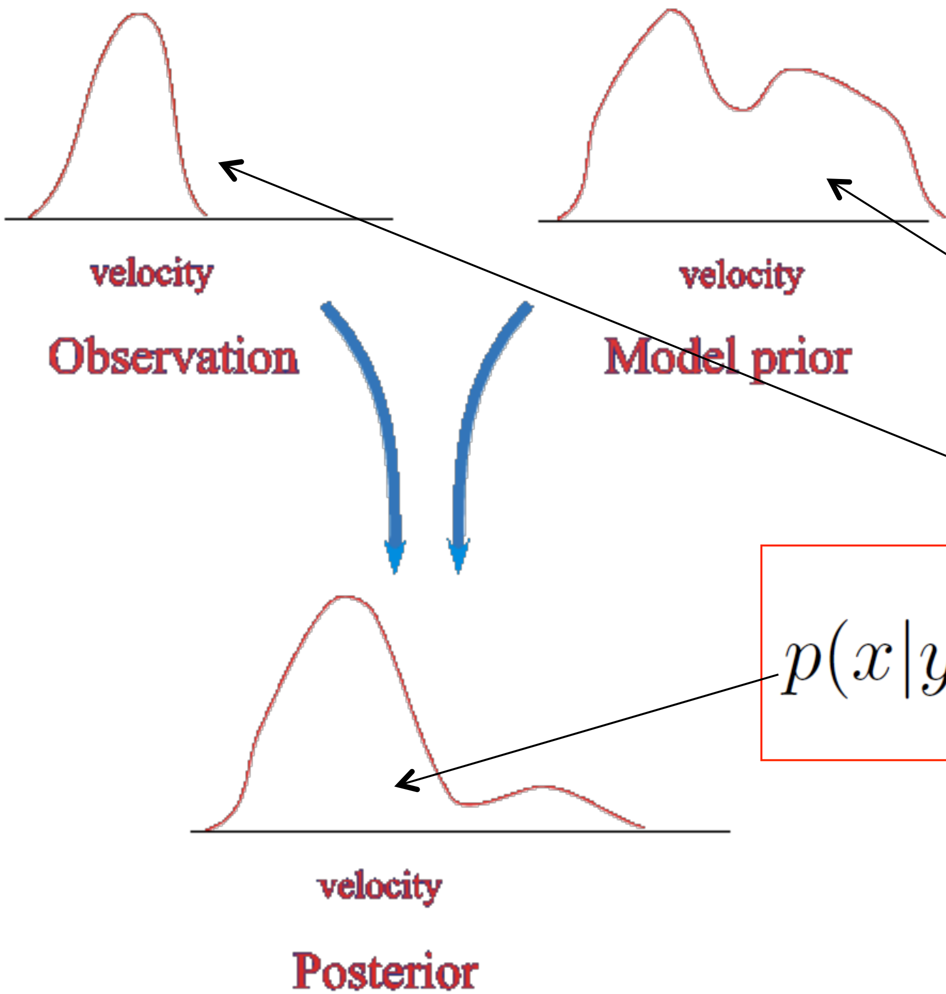
We can use:

$$p(y) = \int p(x, y)\; dx = \int p(y|x)p(x)\; dx$$

Bayes Theorem

$$\boxed{p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\; dx}}$$

Data assimilation: general formulation



Bayes theorem:

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

The solution is a pdf!

No inversion!

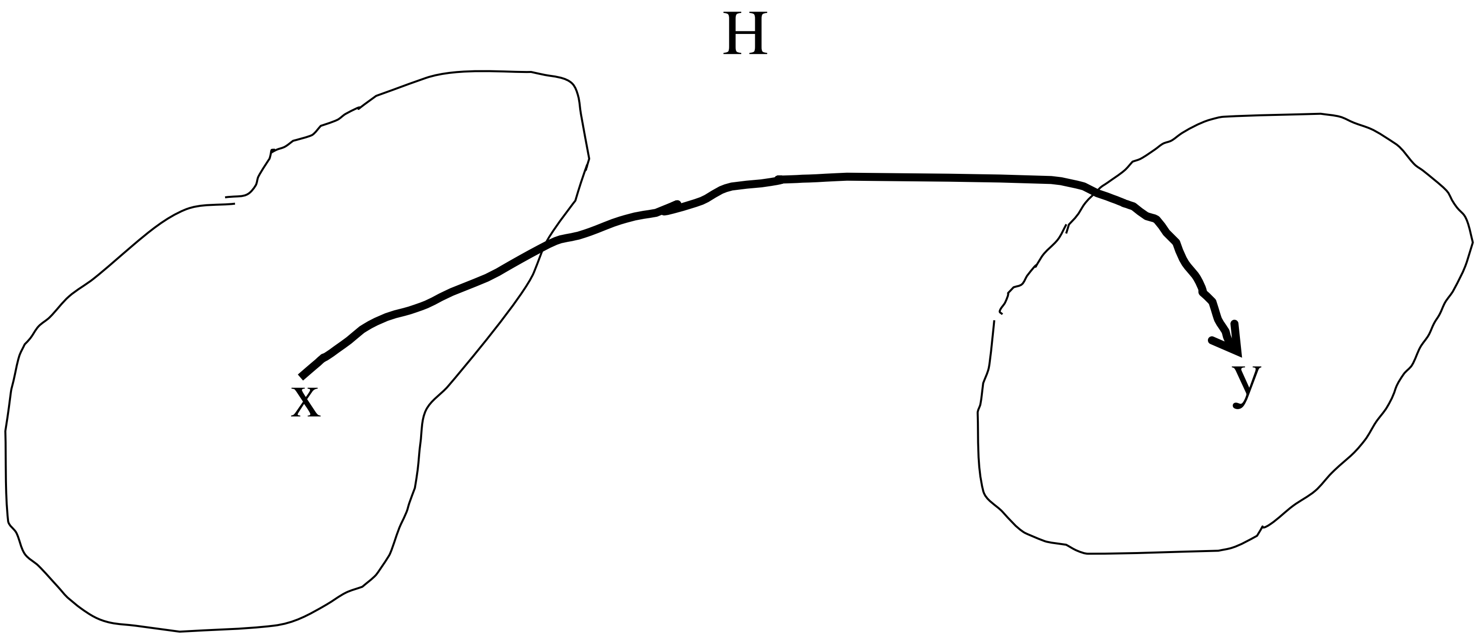
How to use Bayes Theorem

The natural thinking about Bayes Theorem is as follows:

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

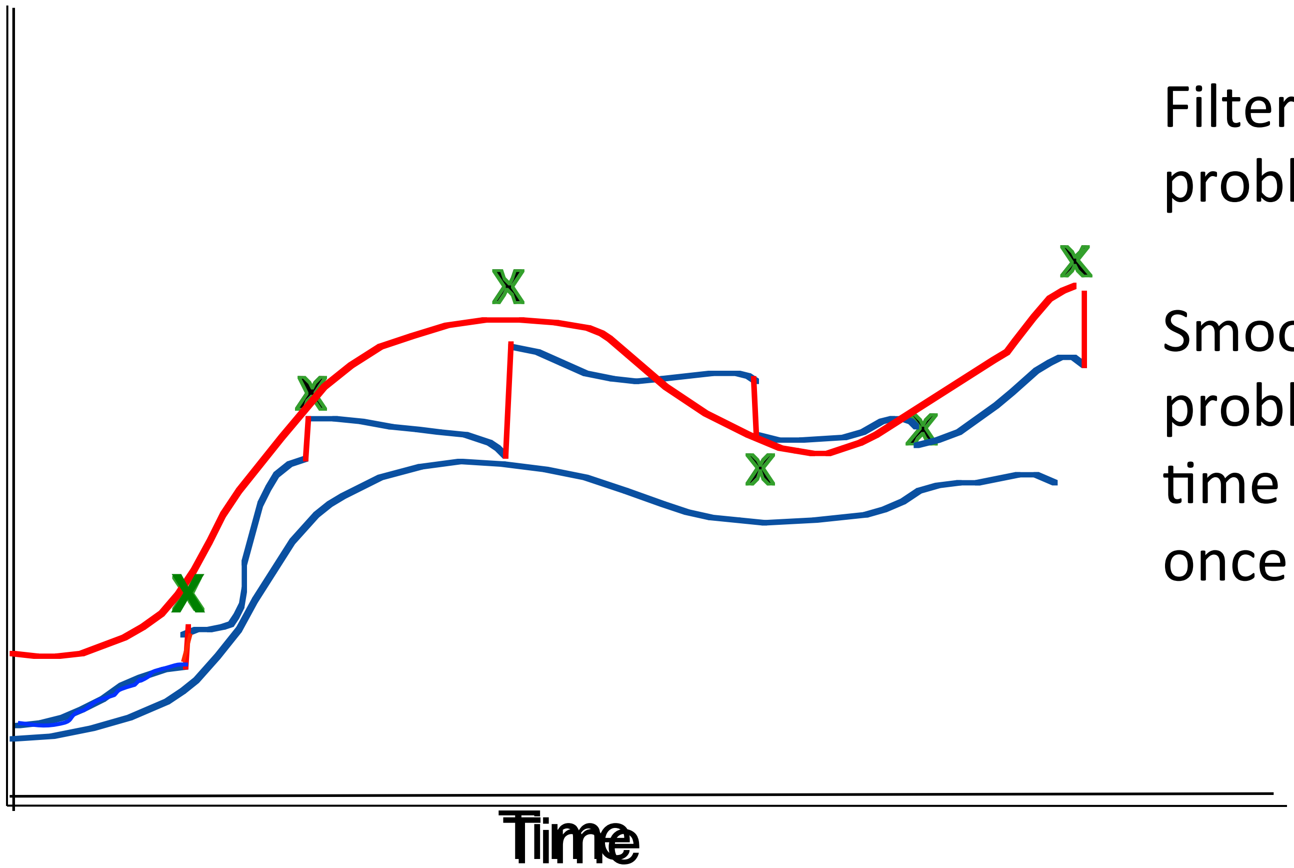
1. Start with the prior, which should contain all information you have about the problem *before looking at the observations*.
2. Multiply the prior with the likelihood to find the posterior.
3. The posterior pdf is the solution to the data-assimilation problem.
4. It is a learning framework.

Observation operator *H*



$$y^n = H(x^n) + \epsilon^n$$

Filters and smoothers



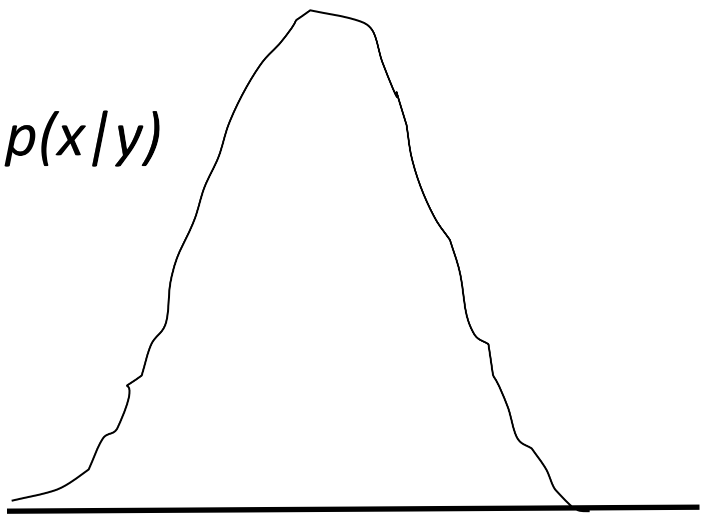
Filter: solve 3D problem sequentially

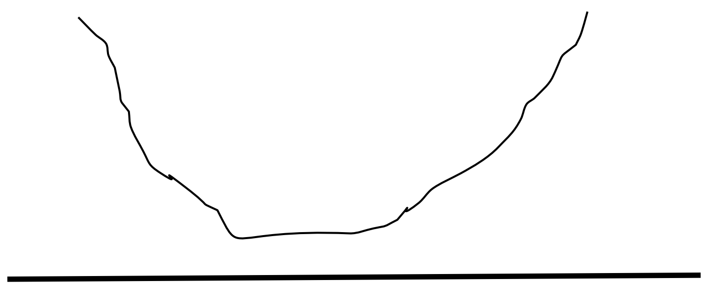
Smoother: solve 4D problem in specific time window all at once

Both can be treated by Bayes Theorem by either defining *x* as a model state state or as a model trajectory.

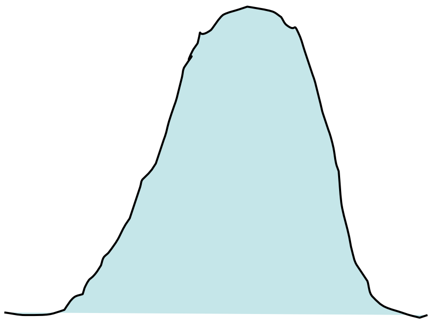
Variational methods

A variational method looks for the most probable state, which is the maximum of this posterior pdf also called the mode.





The Gaussian assumption



$$p(T) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(T-\overline{T})^2}{2\sigma^2}\right]$$

Prior pdf: multivariate Gaussian:

$$\boxed{p(x) \propto \exp\left[-\frac{1}{2}(x-x_b)^T B^{-1}(x-x_b)\right]}$$

Likelihood: multivariate Gaussian

$$\boxed{p(y|x) \propto \exp\left[-\frac{1}{2}(y-H(x))^T R^{-1}(y-H(x))\right]}$$

Variational methods

Instead of looking for the maximum one solves for the minimum of a so-called costfunction.

The pdf can be rewritten as

$$\boxed{p(x|y) \propto \exp\left[-\frac{1}{2}J\right]}$$

in which

$$\boxed{J = (x - x_b)^T B^{-1}(x - x_b) + (y - H(x))^T R^{-1}(y - H(x))}$$

The minimum is found as that state vector for which the derivative is zero. 4DVar needs model adjoint.

Kalman Filter from Bayes Theorem

For the KF we complete the squares to find (only for linear H !!!):

$$p(x|y) \propto \exp\left[-\frac{1}{2}(x-x_a)^T P_a^{-1}(x-x_a)\right]$$

with

$$x_a = x_b + \underbrace{P_b H^T (H P_b H^T + R)^{-1}}_{\text{Kalman Gain } K}(y - H x_b)$$

influence region weighting innovation

$$P_a = (1 - KH)P_b$$

These are the standard Kalman filter equations.

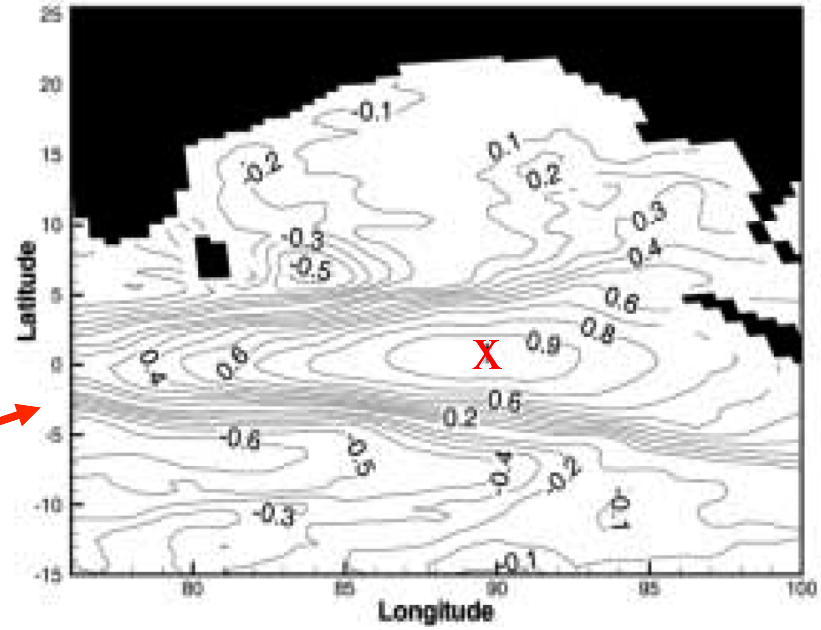
The error covariance:

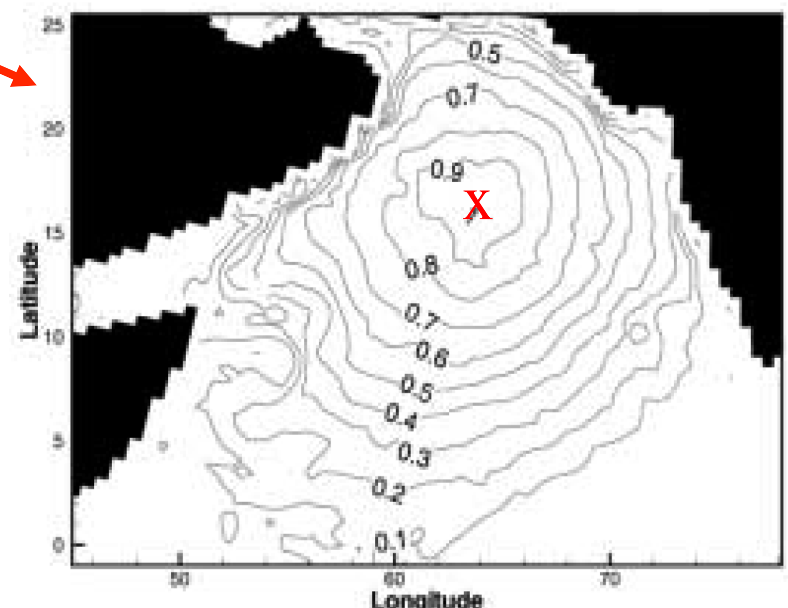
Tells us how model variables co-vary.

Spatial correlation of SSH and SST in the Indian Ocean

In the Kalman filter this comes in via the $P_b H^T$ term:

$$x_a = x_b + P_b H^T (H P_b H^T + R)^{-1} (y - H x_b)$$

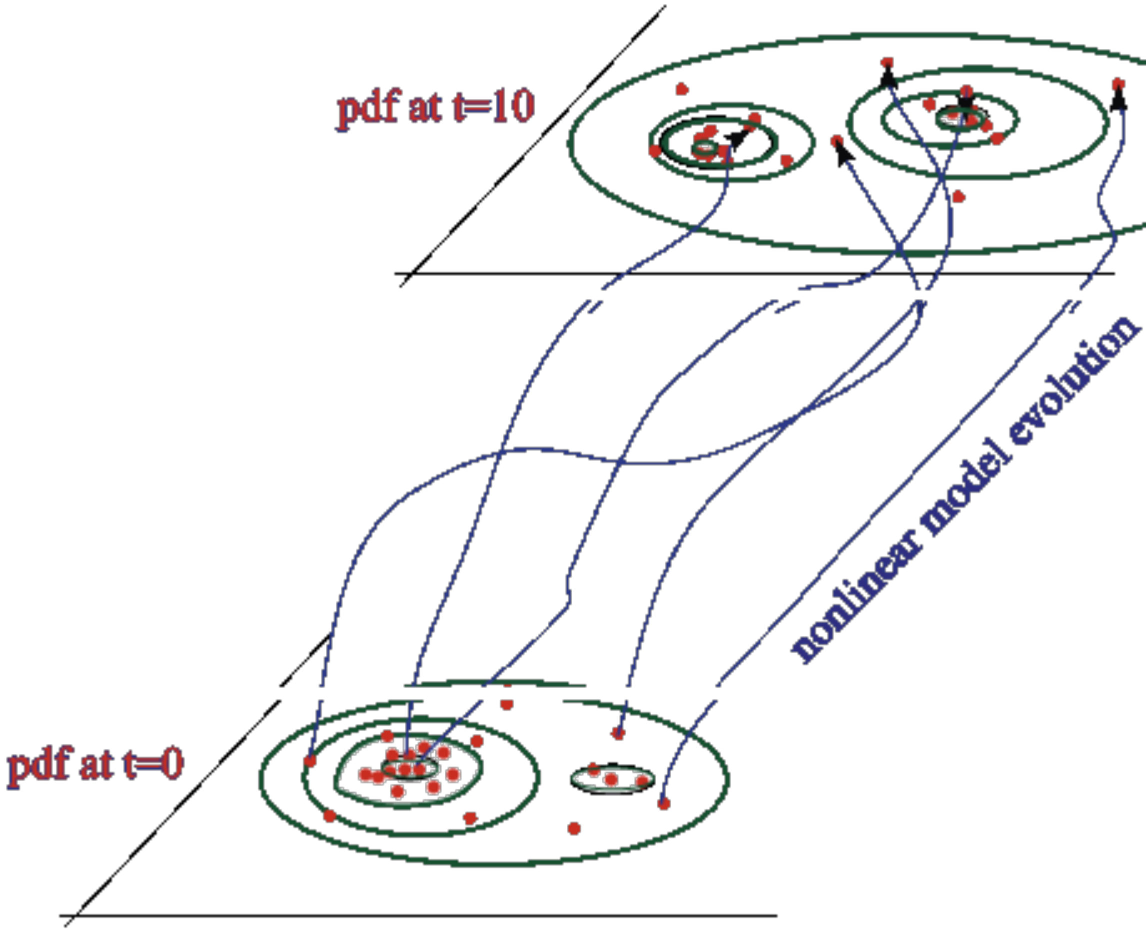




Haugen and Evensen, 2002

Ensemble Kalman Filters

Represent the mean and covariance by an ensemble of model states.



Sampling

Two effects of finite sample size:

- Underestimation of sample covariance.
- Spurious long-range correlations.

Fixes:

- Covariance inflation
- Covariance localization

Localisation observation space

Multiply observation error covariance *R* by a factor that increases as the distance between observation and model grid point increases:

$$R = \rho(d) R$$

in which *d* denotes this distance.

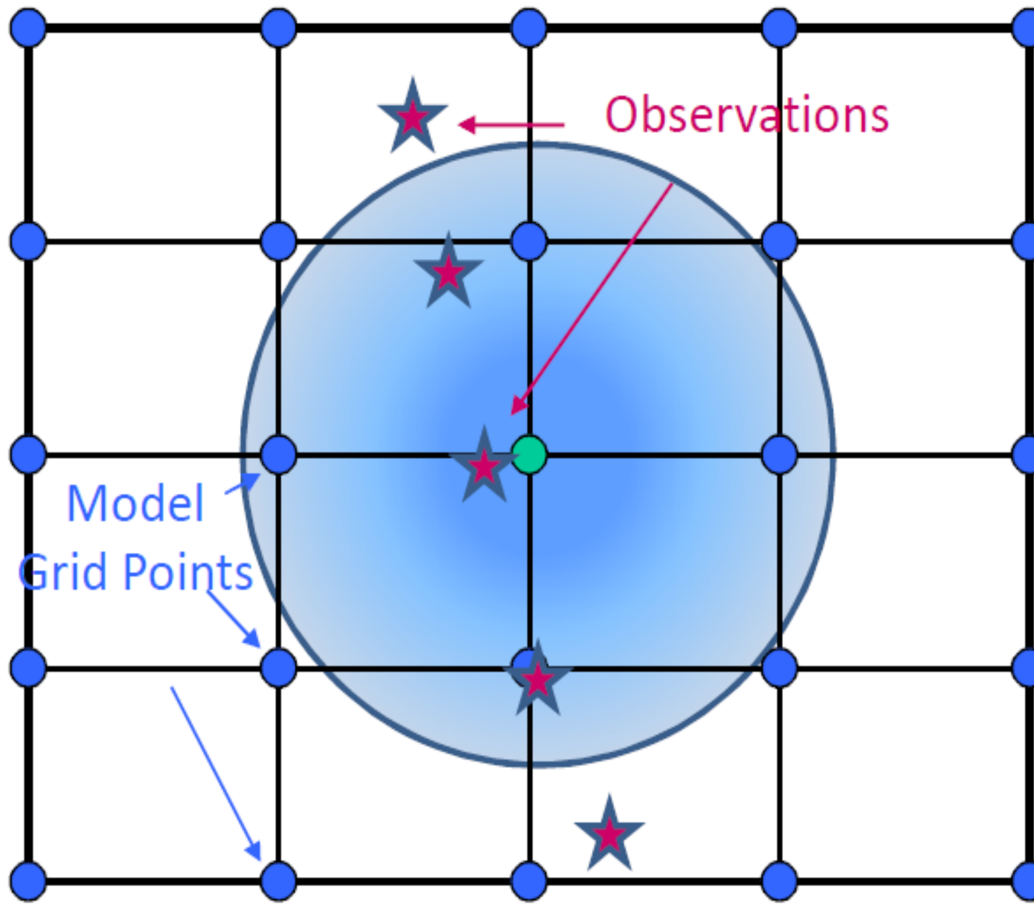


Image courtesy of Steven Greybush.

Advantages of localisation

- Reduces spurious covariances due to small ensemble size
- Decouples ensemble members in different areas --> increase of effective ensemble size
- Brings ‘new blood’ in the ensemble
- Relatively smooth due to covariances

Nonlinear filtering: Particle filter

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

$\Big\downarrow$ Use ensemble

$$\boxed{p(x) = \sum_{i=1}^{N} \frac{1}{N} \delta(x - x_i)}$$

$$\boxed{p(x|y) = \sum_{i=1}^{N} w_i \delta(x - x_i)}$$

with

$$\boxed{w_i = \frac{p(y|x_i)}{\sum_j p(y|x_j)}}$$

the weights.

What are these weights?

- The weight w_i is the normalised value of the pdf of the observations given model state x_i.
- For Gaussian distributed variables is is given by:

$$
\begin{aligned}
w_i &\propto p(y|x_i) \\
&\propto \exp\left[-\frac{1}{2}\left(y - H(x_i)\right)^T R^{-1}\left(y - H(x_i)\right)\right]
\end{aligned}
$$

- This can be calculated directly
- That's all!

No explicit need for state covariances

- 3DVar and 4DVar need a good error covariance of the prior state estimate: complicated
- The performance of Ensemble Kalman filters relies on the quality of the sample covariance, forcing artificial inflation and localisation.
- Particle filter doesn't have this problem, but…

A closer look at the weights

Assume particle 1 is at 0.1 standard deviations *s* of M independent observations.
Assume particle 2 is at 0.2 *s* of the M observations.

The weight of particle 1 will be

$$w_1 \propto \exp\left[-\frac{1}{2}\ (y - H(x_i))\, R^{-1}\, (y - H(x_i))\right] = exp(-0.005M)$$

and particle 2 gives

$$w_2 \propto \exp\left[-\frac{1}{2}\ (y - H(x_i))\, R^{-1}\, (y - H(x_i))\right] = exp(-0.02M)$$

The problem in particle filtering…

The ratio of the weights is

$$\frac{w_2}{w_1} = exp(-0.015M)$$

Take M=1000 to find

$$\boxed{\frac{w_2}{w_1} = exp(-15) \approx 3\ 10^{-7}}$$

Conclusion: the number of independent observations is responsible for the degeneracy in particle filters.

How to make particle filters useful?

1. Introduce ad-hoc localisation to reduce the number of observations in each local area.

2. Use proposal density freedom

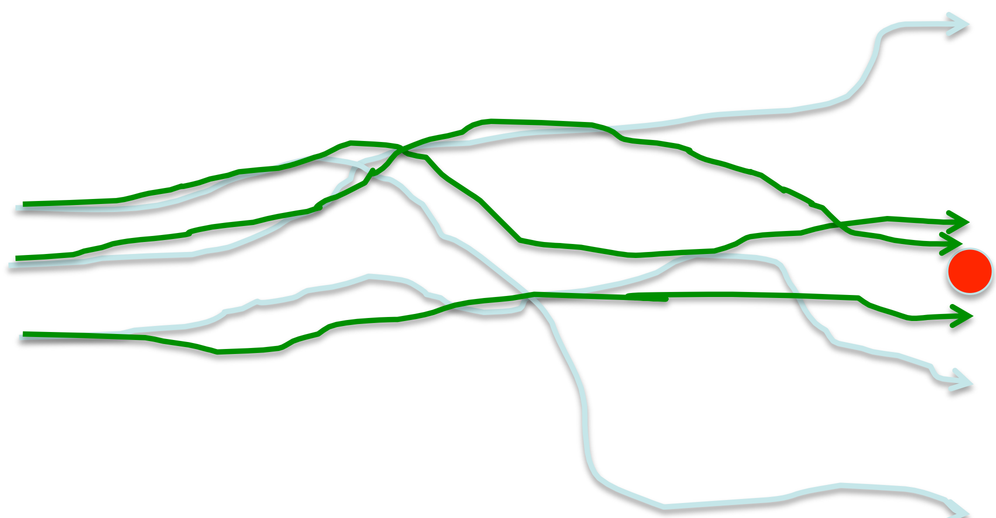
3. Several ad-hoc combinations of Particle Filters and Ensemble Kalman Filters. For instance filters that make sure first 2 moments are correct.

1. Localisation in particle filters

- Easy to make weights spatially varying, similar to observation-space localisation in ETKF.
- Main issue is at the resampling step: how to combine particles from different areas in the domain.
- So need smooth updates without resampling.
- Ensemble Transform Particle Filter (ETPF, Reich, 2014)
- Poterjoy (2014) Complicated scheme that mixes prior and posterior samples and sets minimum weight(!)
- Penny & Myoshi (2016) Too simple.

2. Use proposal density freedom

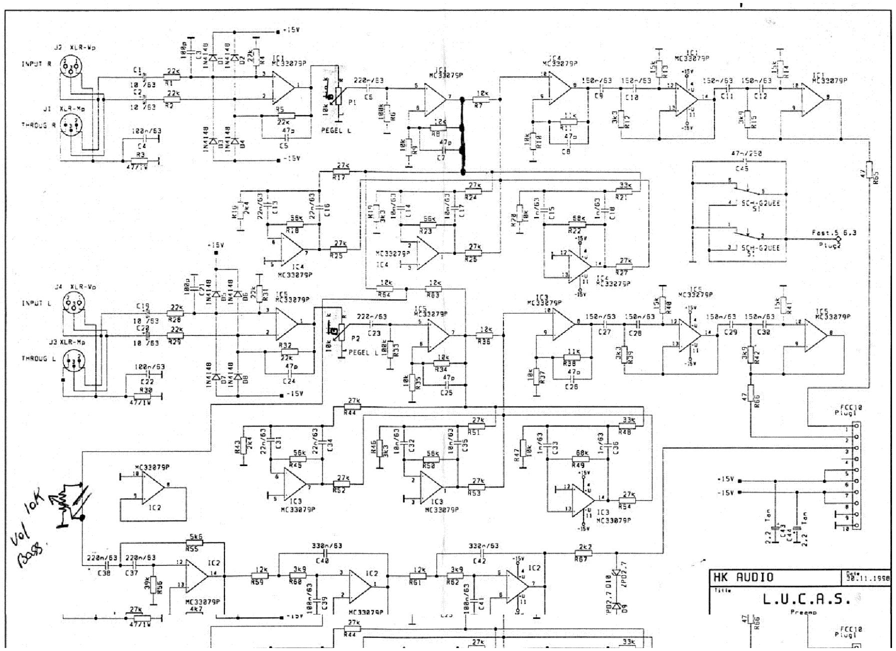
When the ensemble members have reached the observations we are too late: the weights will vary too much.

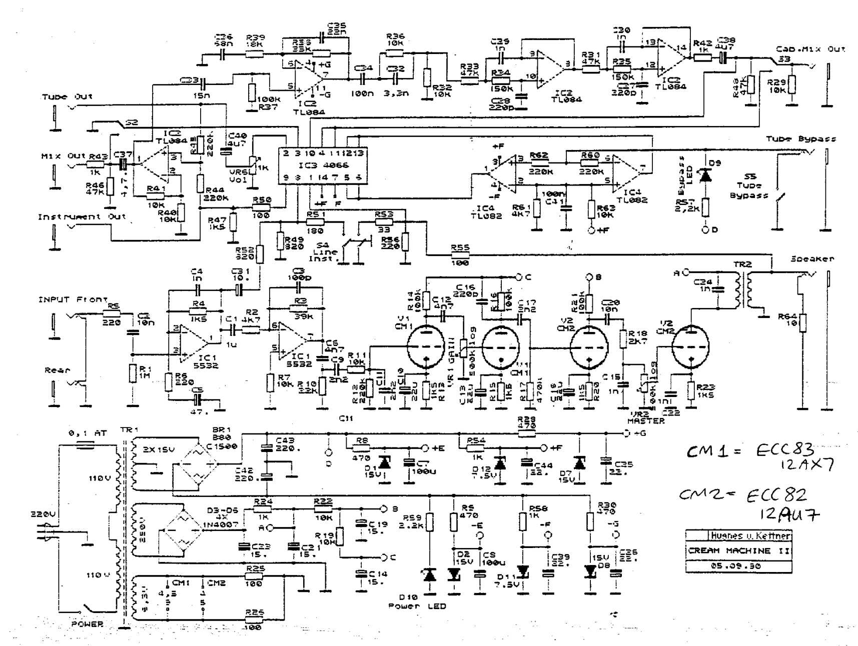


We need to guide them towards the observations ***and*** ensure their weights are equal.

The solution: proposal densities

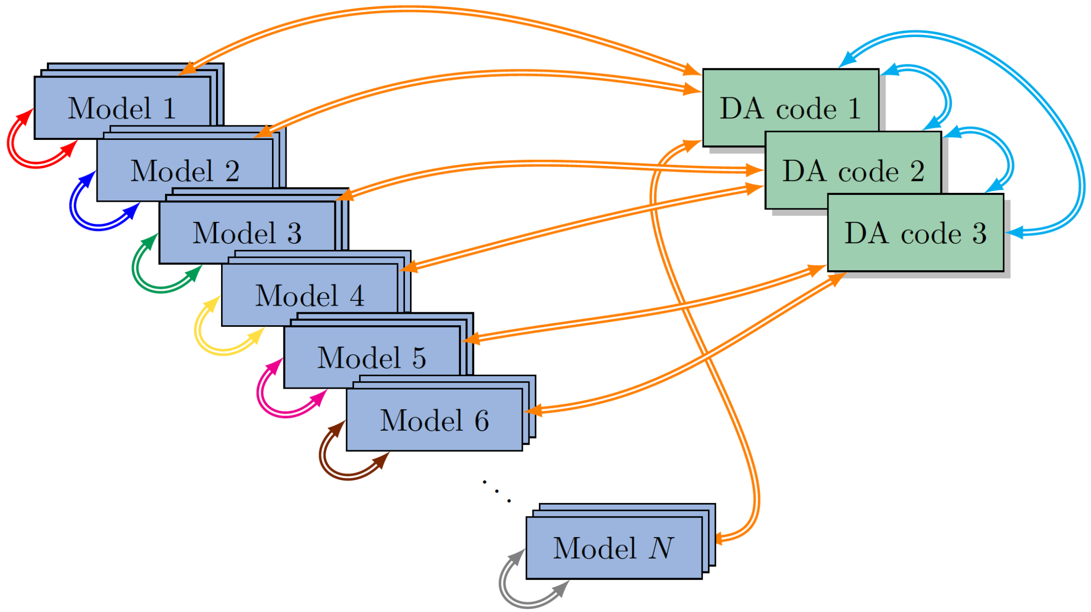
Use a different model and correct in the weights:





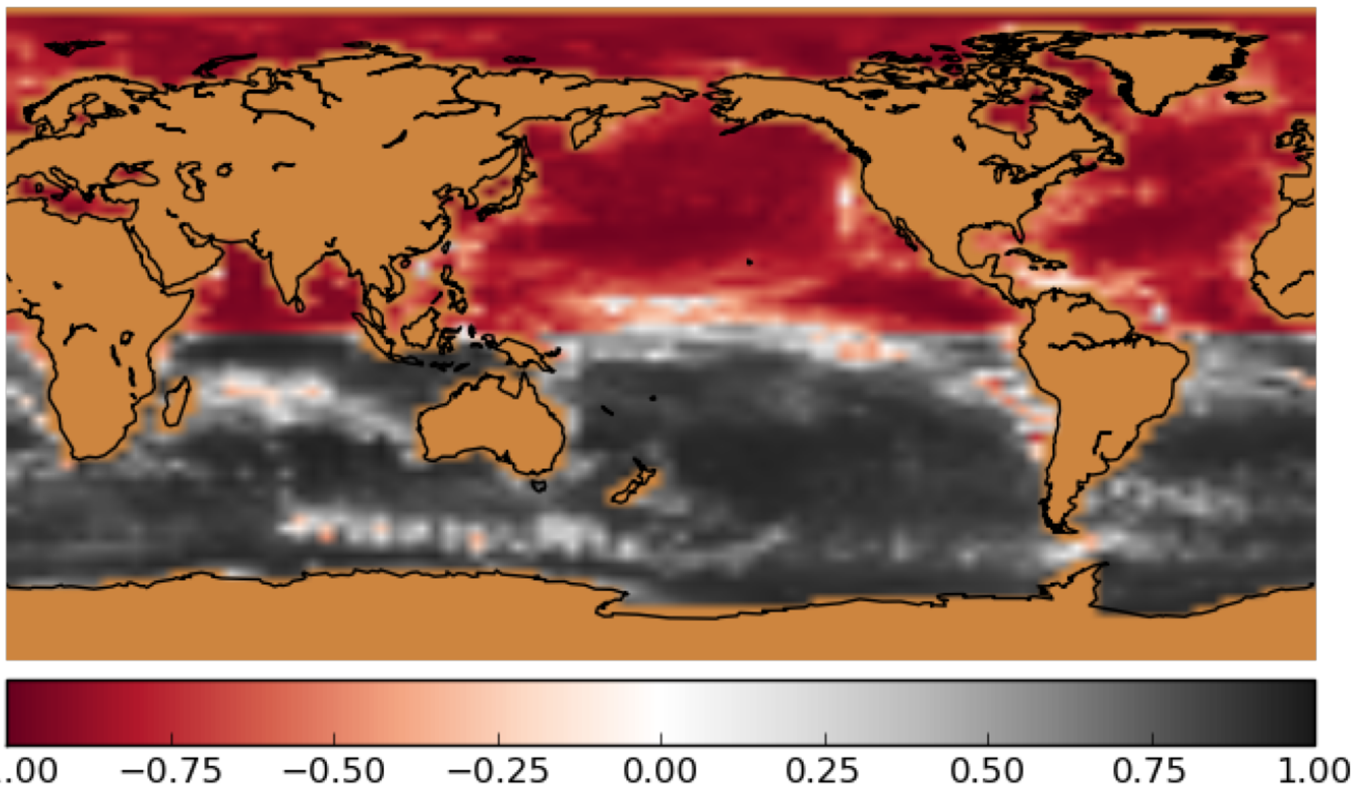
The second model knows about the observations !

EMPIRE data-assimilation framework



Fast coupling of any model to data assimilation codes via MPI, e.g. HadCM3 (2 million), Unified Model (300 million), etc.
Much easier than coupling the model to e.g DART.
See also PDAF.

…and particle filters can be used on e.g. climate models



Correlation atmospheric zonal flow and oceanic meridional flow in HadCM3

What do ensemble members mean?

- They represent the posterior pdf.
- So looking for the best member doesn’t make sense
- A good ensemble is one in which each ensemble member could be replaced with nature,
- So nature and the members are drawn from the same pdf.
- This can be tested, e.g. via rank histograms

Application of DA to paleo climate

- Proper uncertainty estimate is essential!
- Determine error covariance of model state (or evolution)
- Determine error covariance of observations.
- Use Bayes Theorem to get posterior pdf (or best estimate plus error estimate)
- Use Bayesian framework to add new observations, it is a learning framework

How to find error covariance of model state?

- Variances from knowledge of model behaviour, comparison with observations (past and present)
- Correlations via knowledge of the system (e.g. physical relation between variables such as geostrophic balance)
- Or generate first estimate from ensemble of snap shots from long model run. This gives the total variance in the climate model, actual error covariance can be taken as fraction of this.
- Or generate ensemble of forecasts with different lead times, e.g. 1 year and 2 years.

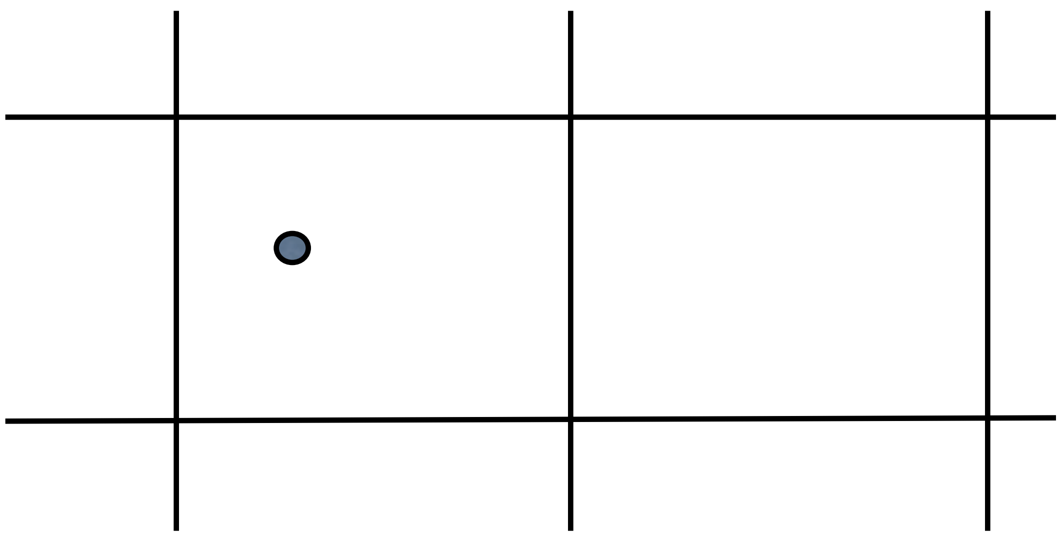
How to find error covariance of observations?

- Instrument errors
- Representation errors

What are representation errors?

Extra uncertainty that arises in data assimilation because model and observations have different representation of reality:

- Different resolution (weather forecasting)
- Isotope ratio versus model temperature
- Etc..



How do they arise in DA?

Start from Bayes Theorem:

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

Representation differences between model and observations, so representation errors arise in the likelihood, not in the prior!

Likelihood with unknown *H*

The observation equation reads $y = H(x) + \epsilon$
but *H* is not known exactly.
Use

$$p(y|x) = \int p(y, H|x)\; dH$$

so

$$p(y|x) = \int p(y|H, x)p(H|x)\; dH$$

$p(y|H, x)$ is the normal likelihood with fixed *H*

$p(H|x)$ is the pdf that describes the uncertainty in *H*

$p(y|x)$ is the convolution of the two.

Gaussian example

Assume *H* is given by: $H(x) = \tilde{H}(x) + \eta$

so $$y = H(x) + \epsilon = \tilde{H}(x) + \eta + \epsilon$$

Gaussian instrument errors:

$$p(y|H,x) \propto \exp[-1/2(y - \tilde{H}(x) - \eta)^T R^{-1} (y - \tilde{H}(x) - \eta)]$$

Gaussian observation operator errors:

$$p(H|x) \propto \exp[-1/2\ \eta^T C^{-1} \eta]$$

use in

$$p(y|x) = \int p(y|H,x) p(H|x)\, dH$$

to find

$$p(y|H,x) \propto \exp[-1/2(y - \tilde{H}(x))^T (R + C)^{-1} (y - \tilde{H}(x))]$$

Likelihood when physics is missing

Introduce $$z = (x, \tilde{z})^T$$

and use

$$\boxed{p(y|x) = \int p(y|x, \tilde{z})p(\tilde{z}|x)d\tilde{z} = \int p(y|z)p(\tilde{z}|x)d\tilde{z}}$$

$p(y|z)$ is the instrument error

$p(\tilde{z}|x)$ is the representation error

$p(y|x)$ is the convolution of the two

Is observation sparseness or poor accuracy a problem?

- In principle: NO!
- Estimate errors as good as possible
- Explain very clearly how you did that
- Accept that if prior errors are large and observation errors are large then also the posterior errors will be large
- Use Bayian framework so that new observations can always be added; no need for using old observations again.
- In practise: ok, not that nice…

Conclusions

- Data assimilation is based on a solid mathematical framework: Bayes Theorem.
- Need to provide error covariances of observations
- Need to provide error covariances of the prior, used at every time window in variational methods
- For ensemble Kalman Filters prior is only needed at start of reanalysis, but inflation and localisation are essential at each observation time
- Variational-EnKF hybrids are popular (no adjoint needed)
- Particle filters don't need state covariance matrices explicitly
- Efficient particle filters need localisation and inflation, or model error covariance Q
- Methods should work for paleo climatology too…