Data assimilation in biogeochemistry: Adapting the paradigm of numerical weather prediction

Peter Rayner
University of Melbourne

Outline of series

1. Basic approach with some simple examples;

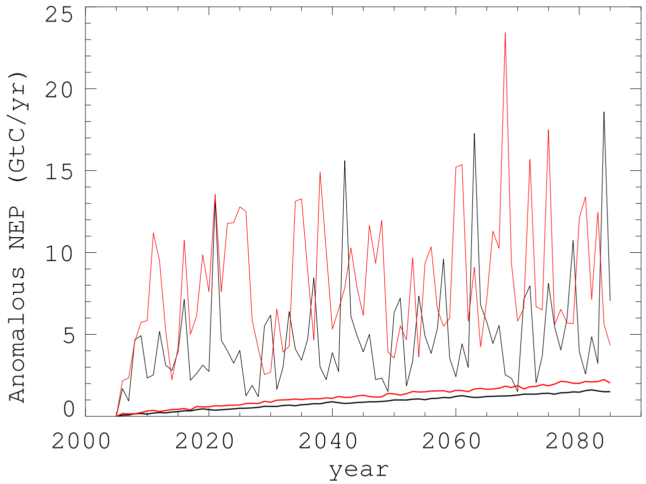
2. What can go wrong and how would we know?

3. Some advanced uses, model development and evaluation.

Outline for Lecture One

- Motivation: An example of data assimilation for climate;
- The minefield of nomenclature and notation;
- Data assimilation as Bayesian inference;
- Some simple examples;
- Looking hard at each component.

Motivation



Uncertainty in terrestrial uptake, 2000–2090. Black lines = current climate, red = climate change. Thin lines = original model, thick = after data.

- Rayner et al., Phil. Trans. 2010;
- Uncertainties completely dominated by climate change;
- Greatly reduced by confronting with data.

The problem

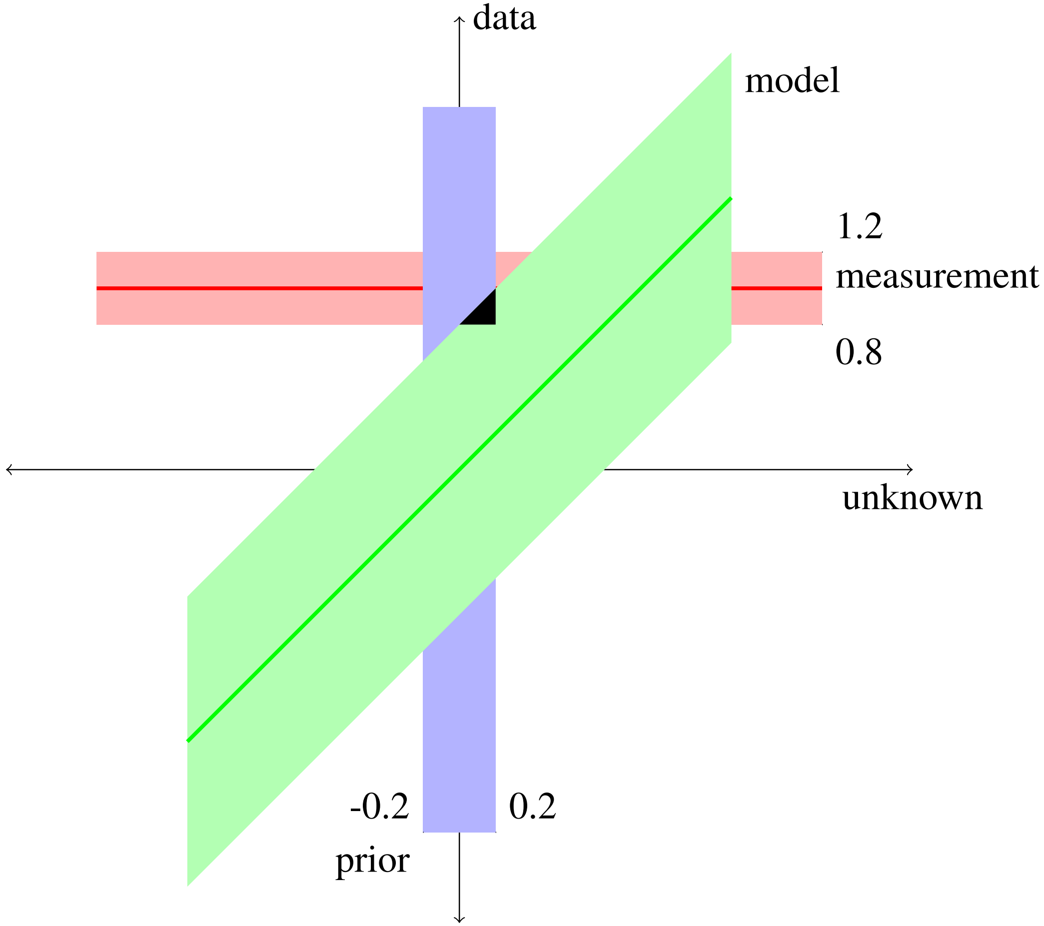
- To improve our knowledge of the state and functioning of a physical system given some observations.

- “State” means the value of physical quantities which may evolve, usually the variables in a numerical model;

- “Function” means the fixed values or even functional forms of the laws governing the system.

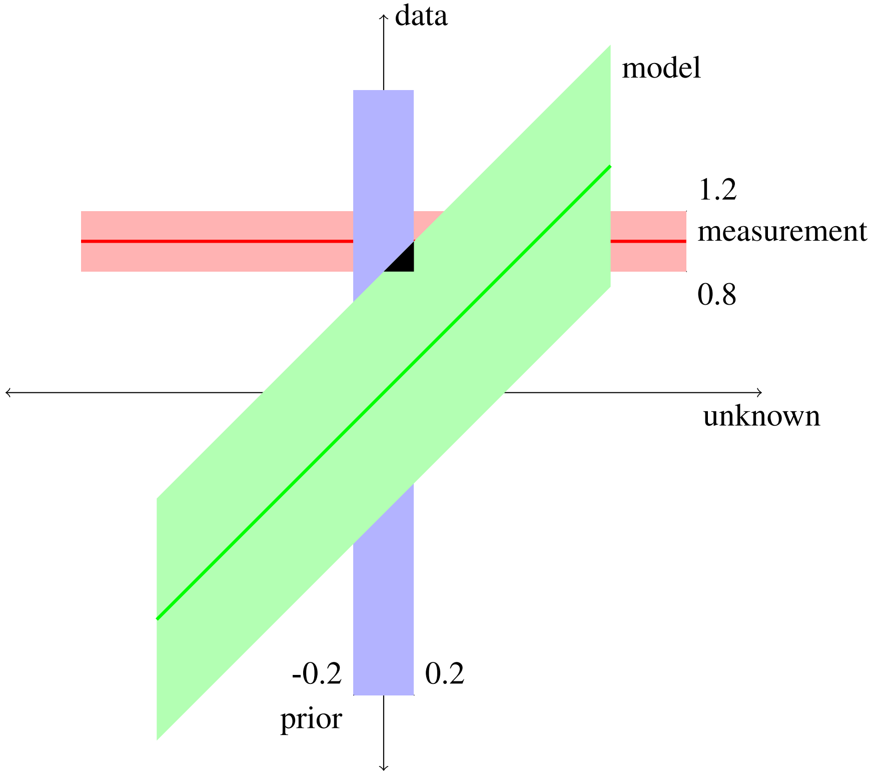
Name	Symbol	Description	Examples
Parameters	$\vec{p}$	Quantities not changed by model	ξ (buffer factor), b_a (terrestrial flux amplitude)
State variables	$\vec{v}$	Quantities altered by model	leaf area, DIC
Unknowns[1]	$\vec{x}$	Quantities exposed to optimisation	ξ, $c_I(t=0)$
Observables	$\vec{o}$	Measurable quantities, may be in $\vec{v}$	c_A, total carbon
Observation operator		Transforms $\vec{v}$ to $\vec{o}$	1, $c_I + c_O$
Model	$\mathbf{M}$	Predicts $\vec{v}$ given $\vec{p}$ and $\vec{v}(t=0)$	
Data	$\vec{d}$	Measured values of $\vec{o}$	

Data Assimilation in One Picture

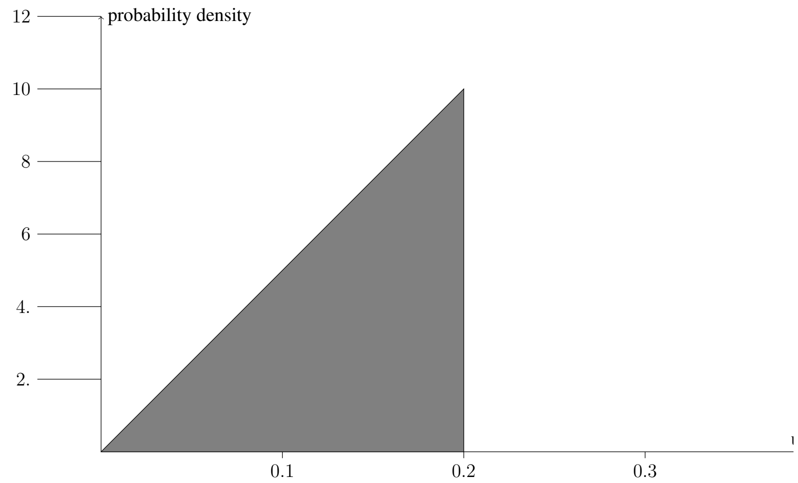


- Unknown on X-axis, obs on Y-axis;
- Light-blue = prior unknown
- Light-red = obs
- Green = model;
- Black = solution.

Well, almost one picture



Solution is multiplication of input PDFs.

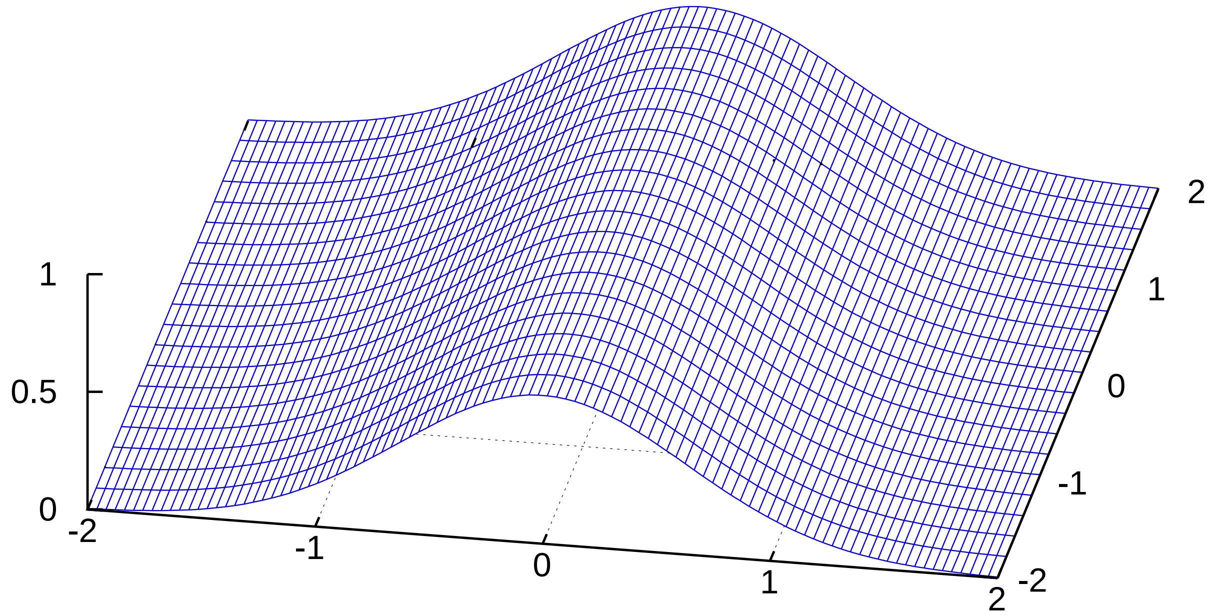


Final PDF projects triangle onto "unknown" axis.

Notes

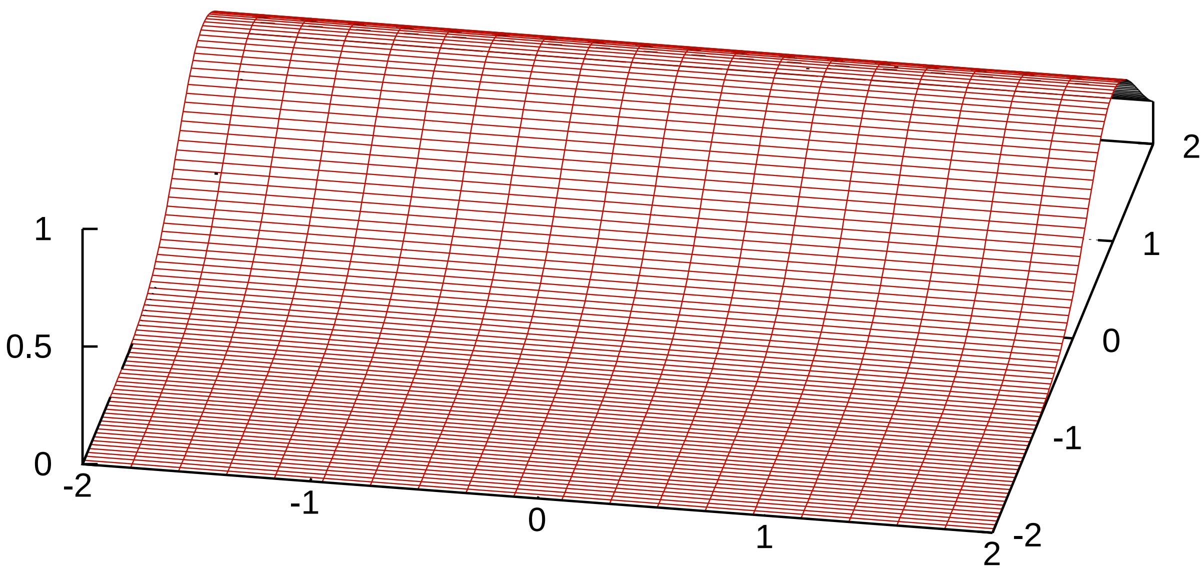
- Solution is multiplication of PDFs;
- Solution can be constructed with only forward models;
- Normalization doesn't usually matter.

Gaussian Prior



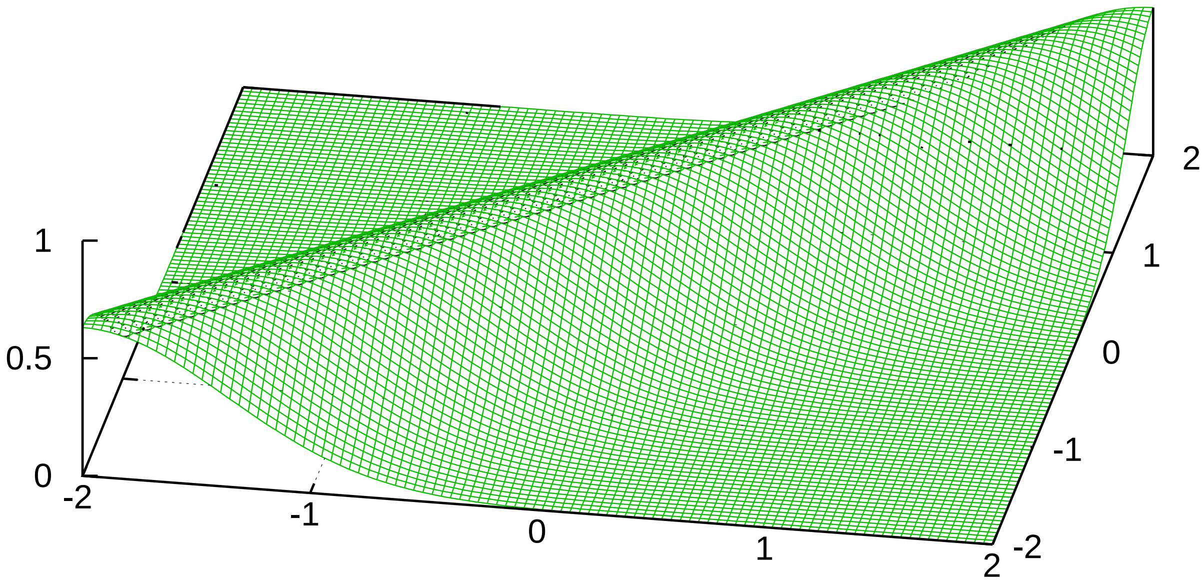
$$\frac{1}{\sqrt{2\pi}\sigma_P} \exp -\frac{x^2}{2\sigma_P^2}$$

Data



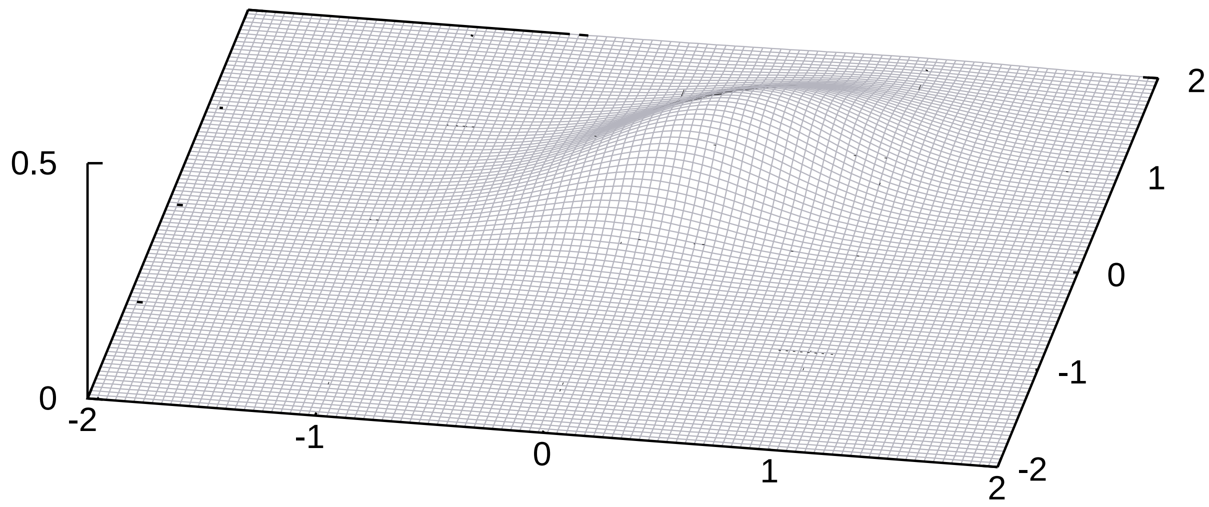
$$\frac{1}{\sqrt{2\pi}\sigma_D}\exp -\frac{(y-1)^2}{2\sigma_D^2}$$

Model



$$\frac{1}{\sqrt{2\pi}\sigma_M}\exp -\frac{(y-M(x))^2}{2\sigma_M^2}$$

Prior plus Data plus Model



$$\frac{1}{\sqrt{2\pi}\sigma_P\sigma_D\sigma_M}\exp -\frac{x^2}{2\sigma_P^2} \times \exp -\frac{(y-1)^2}{2\sigma_D^2} \times \exp -\frac{(y-M(x))^2}{2\sigma_M^2}$$

"Solving" the Inverse Problem

- The joint PDF *is* the solution;
- For Gaussians the solution can be represented by a mean and variance;
- These can be misleading.

A simple example

$$P(x,y) = \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y\sigma_M} \exp - \frac{(x-x_0)^2}{2\sigma_x^2} \times \exp - \frac{(y-D)^2}{2\sigma_y^2} \times \exp - \frac{(y-M(x))^2}{2\sigma_M^2}$$

- $x_0 = 0$, $D = 1$, $M = 1$, $\sigma_x = \sigma_y = \sigma_M = 1$;
- Multiplying exponentials $\leftrightarrow$ adding exponents;

$$P(x,y) = \frac{1}{\sqrt{2\pi}} \exp - \left[\frac{x^2}{2} + \frac{(y-1)^2}{2} + \frac{(y-x)^2}{2} \right]$$

Solution Continued

$$P(x,y) = \frac{1}{\sqrt{2\pi}} \exp - \left[\frac{x^2}{2} + \frac{(y-1)^2}{2} + \frac{(y-x)^2}{2} \right]$$

- Finding most likely value means maximizing probability

- Maximizing negative exponential means *minimizing*:

$$J = \frac{1}{2} \left[x^2 + (y-1)^2 + (y-x)^2 \right]$$

- Example of least squares cost function.

Solution Continued

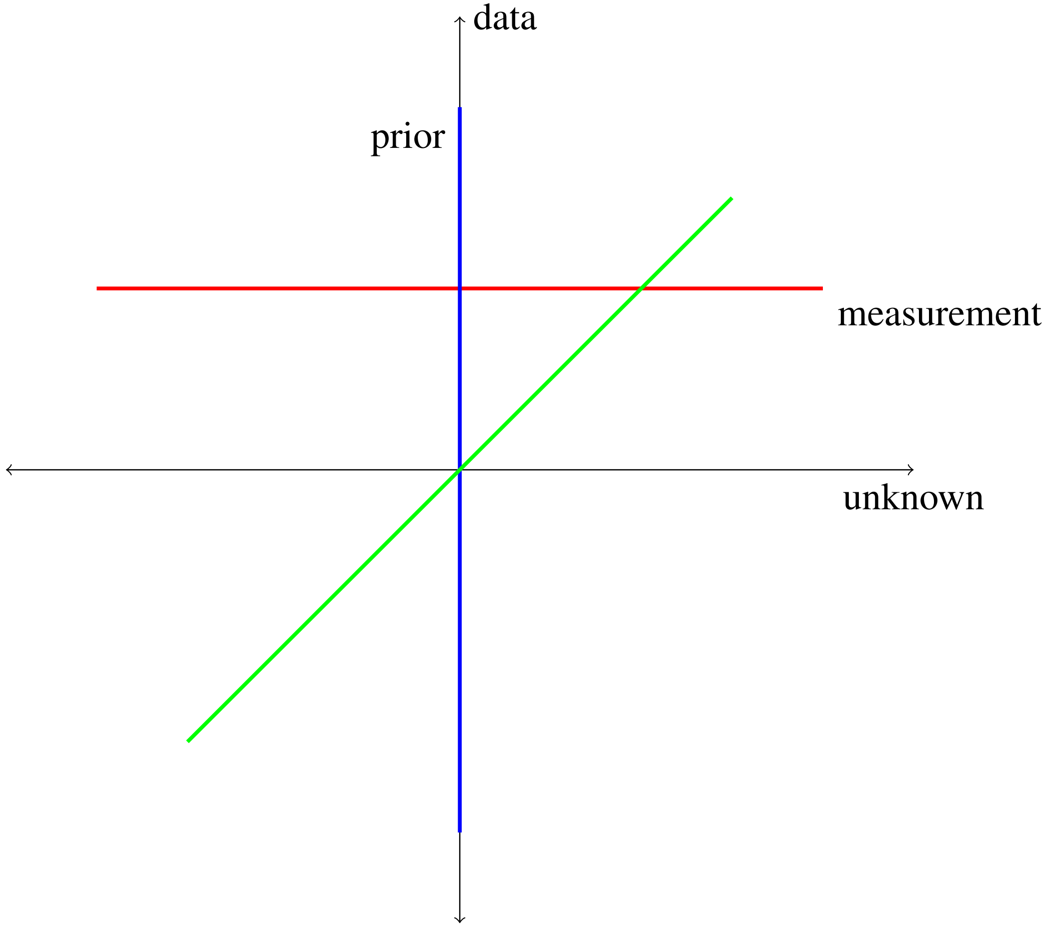
$$J = \frac{1}{2}\left[x^2 + (y-1)^2 + (y-x)^2\right]$$

- To maximize set $\frac{\partial J}{\partial x} = 0$ and $\frac{\partial J}{\partial y} = 0$

$$\begin{aligned} 2x - y &= 0 && (1) \\ 2y - x - 1 &= 0 && (2) \end{aligned}$$

- $x = \frac{1}{3}$, $y = \frac{2}{3}$

Illustrating Solution



- Prior estimate is intersection of red and blue lines (0, 1).
- Solution is pulled directly towards model;
- Solution is compromise between prior, measurement and model;
- Solution depends on both values and uncertainties.

More detail on Uncertainties

- Prior PDF is distribution of true value *deliberately ignoring* measurements we intend to use. Often expressed as distribution around value but not necessary.

- PDF of data is distribution of true value, usually distributed around a measurement;

- PDF of model describes distribution of true value given particular value of "unknown". Almost never available.

First Simplification

- Often we are not interested in estimating the observable;
- For Gaussian PDFs we can pretend our model is perfect and add observational and modelling error *variances* (Tarantola 2004, P202);
- Thus

$$J = \frac{1}{2}\left[x^2 + (y-1)^2 + (y-x)^2\right]$$

  becomes

$$J^* = \frac{1}{2}\left[x^2 + (x-1)^2/2\right]$$

- Yields $x = \frac{1}{3}$ but *not* $y = \frac{2}{3}$.

Recursive estimation

- Multiplication of PDFs can be done in any order and many at a time or singly;
- If we preserve the full PDF we can include observations as they arrive;
- For Gaussians PDF described by means and variances;
- Information is always added so that PDFs are always refined.

Batch and Sequential Methods

BATCH

- Handle all obs at once;
- PDFs for priors and obs unrestricted;
- Model error hard to include;
- Classic example 4dVar for weather prediction.

SEQUENTIAL

- Handle obs as they arrive;
- PDFs for obs restricted (time correlations hard);
- Model error handled very naturally;
- Kalman Filter.

A few Example Applications

- What are the unknowns?
- What is the prior estimate?
- What are the observations?
- What is the model?
- How do they handle the time domain?

Numerical Weather Prediction 4dVar

- Unknown is 3d grid of atmospheric variables at fixed time;
- Prior is previous forecast;
- Observations include in situ and satellite measurements over a fixed time window;
- Model combines dynamic evolution of atmosphere with observation operators;
- All observations handled at once;
- doesn't *usually* have explicit model error.

Numerical Weather Prediction, Kalman filtering

- Unknown is 3d grid of atmospheric variables at *each* time;
- Prior is previous posterior;
- Observations include in situ and satellite measurements within one timestep;
- Dynamic model and observation operators separated;
- *Always* has explicit model error.

Atmospheric Flux Inversion

- Unknown is space-time distribution of surface fluxes;
- Prior often comes from biogeochemical model;
- Observations are atmospheric concentration;
- Model is atmospheric transport;
- All observations *usually* handled at once;
- Model error sometimes handled via model ensemble.

Biogeochemical data assimilation

- Confusing terminology;
- Unknowns are parameters in model;
- Priors from independent experiment or literature;
- Many different observations (fluxes, concentrations, vegetation indices, ocean colour etc);
- Dynamic model and obs operators separated;
- Equally split between batch and sequential.

Linear Gaussian Case

- Unknowns and data are vectors $\vec{x}$ and $\vec{d}$;
- σ^2 replaced with variance/covariance matrices $\mathbf{C}$ for $\vec{x}$ and $\vec{d}$;
- Model M becomes matrix $\mathbf{M}$;
- Use usual simplification of assuming perfect model and adding data and model uncertainties.

Solution

$$P(\vec{x}) = K \frac{1}{\sqrt{\det \mathbf{C}(\vec{x}_0) \det \mathbf{C}(\vec{y})}} \exp -\frac{1}{2}(\vec{x} - \vec{x}_0)^T \mathbf{C}^{-1}(\vec{x}_0)(\vec{x} - \vec{x}_0) \exp -\frac{1}{2}(\mathbf{M}\vec{x} - \vec{y})^T \mathbf{C}$$

Minimize

$$J = (\vec{x} - \vec{x}_0)^T \mathbf{C}^{-1}(\vec{x}_0)(\vec{x} - \vec{x}_0) + (\mathbf{M}\vec{x} - \vec{y})^T \mathbf{C}^{-1}(\vec{y})(\mathbf{M}\vec{x} - \vec{y})$$

Continued

$$J = (\vec{x} - \vec{x}_0)^T \mathbf{C}^{-1}(\vec{x}_0)(\vec{x} - \vec{x}_0) + (\mathbf{M}\vec{x} - \vec{y})^T \mathbf{C}^{-1}(\vec{y})(\mathbf{M}\vec{x} - \vec{y})$$

Yields

$$\vec{x} = \vec{x}_0 + \mathbf{C}(\vec{x}_0)\mathbf{M}^T \left[\mathbf{M}\mathbf{C}(\vec{x}_0)\mathbf{M}^T + \mathbf{C}(\vec{y})\right]^{-1} (\vec{y} - \mathbf{M}\vec{x}_0)$$

$$\mathbf{C}^{-1}(\vec{x}) = \mathbf{C}^{-1}(\vec{x}_0) + \mathbf{M}^T \mathbf{C}^{-1}(\vec{y})\mathbf{M}$$

Summary

- Data assimilation is an example of Bayesian Inference;
- BI itself follows from rules for combining PDFs;
- Techniques like least squares minimisation are special cases for particular types of PDF;
- Most approaches such as Kalman Filtering and 4dVar can be expressed with this formalism.