Data Assimilation 1

Alan O'Neill
National Centre for Earth Observation
UK

Plan

- Motivation & basic ideas
- Univariate (scalar) data assimilation
- Multivariate (vector) data assimilation
  - 3d-Variational Method (& optimal interpolation)
  - Kalman Filter (+ extended KF)
  - Ensemble methods (+ particle filter)
  - 4d-Variational Method
- Applications of data assimilation in earth system science

What is data assimilation?

Data assimilation is the technique whereby observational data are combined with output from a numerical model to produce an optimal estimate of the evolving state of the system.

What are the benefits of data assimilation?

- Quality control
- Combination of data
- Errors in data and in model
- Filling in data poor regions
- Designing observing systems
- Maintaining consistency
- Estimating unobserved quantities
- Parameter estimation in models *****

The Data Assimilation Problem

How can we combine noisy measurements of a system with output from an imperfect numerical model to get the best estimate of the (evolving) state of the system?

Answer:
Use Bayes' Theorem with the following information:

- The observations
- Their errors
- Predictions by a numerical model of the system
- The errors in these predictions

The key idea is to combine observations with predictions giving more weight to information with the least error. But errors may not be well known! Internal consistency checks on our state estimates are possible, but also need independent (unassimilated data).

Conditional Probability & Bayes' Theorem

$p(A,B) = p(A \mid B)\,p(B) = p(B \mid A)\,p(A),$

where A and B are two random events

Bayes' Theorem: $$p(A \mid B) = \frac{p(B \mid A)\,p(A)}{p(B)}$$

$$\backslash \quad p(x \mid y) = \frac{p(y \mid x)\,p(x)}{p(y)},$$

where x is a state variable of the system we wish to estimate, and z is a measurement of that variable.

So if we have some prior information about $p(x)$, we can update that information with an observation y to get $p(x \mid y)$, the probability that the system variable has value x given that a measurement z of that variable has been made. We call it the posterior pdf.

Errrors everywhere

All significant sources of uncertainty should be accounted for in data assimilation

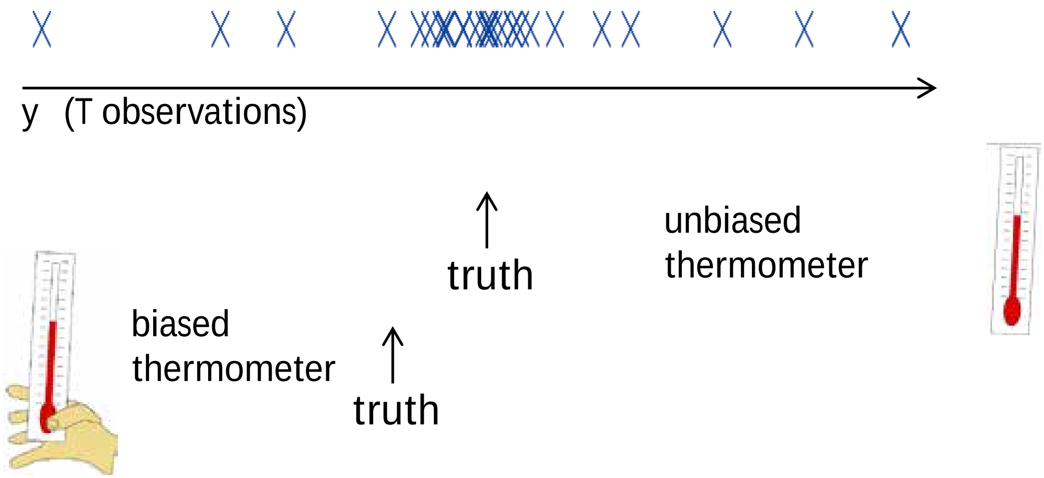
Random errors:

- background (a-priori) errors
- observation errors
- model errors
- representivity errors

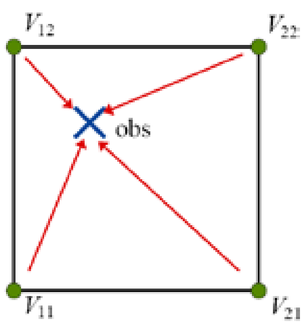
Systematic errors:

- biases in background
- biases in observations
- biases in model

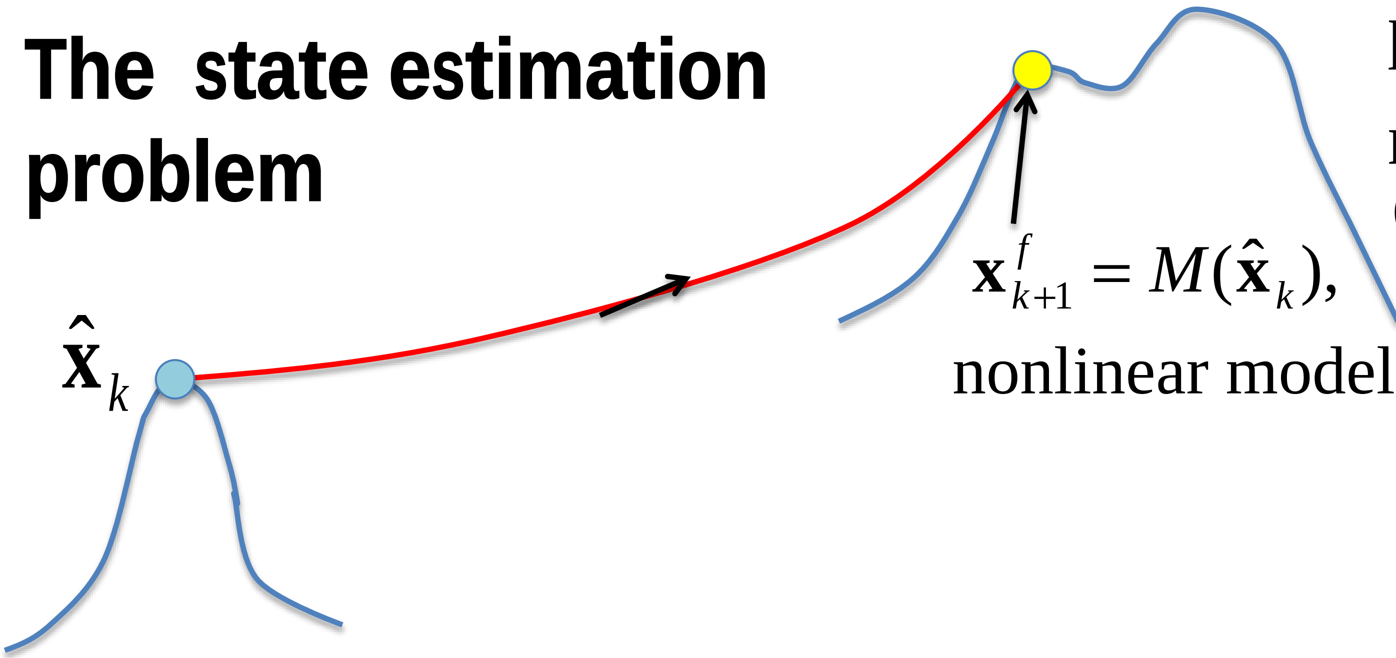
Example 1 – repeated observations of air temperature



Example 2 – representivity errors due to model grid

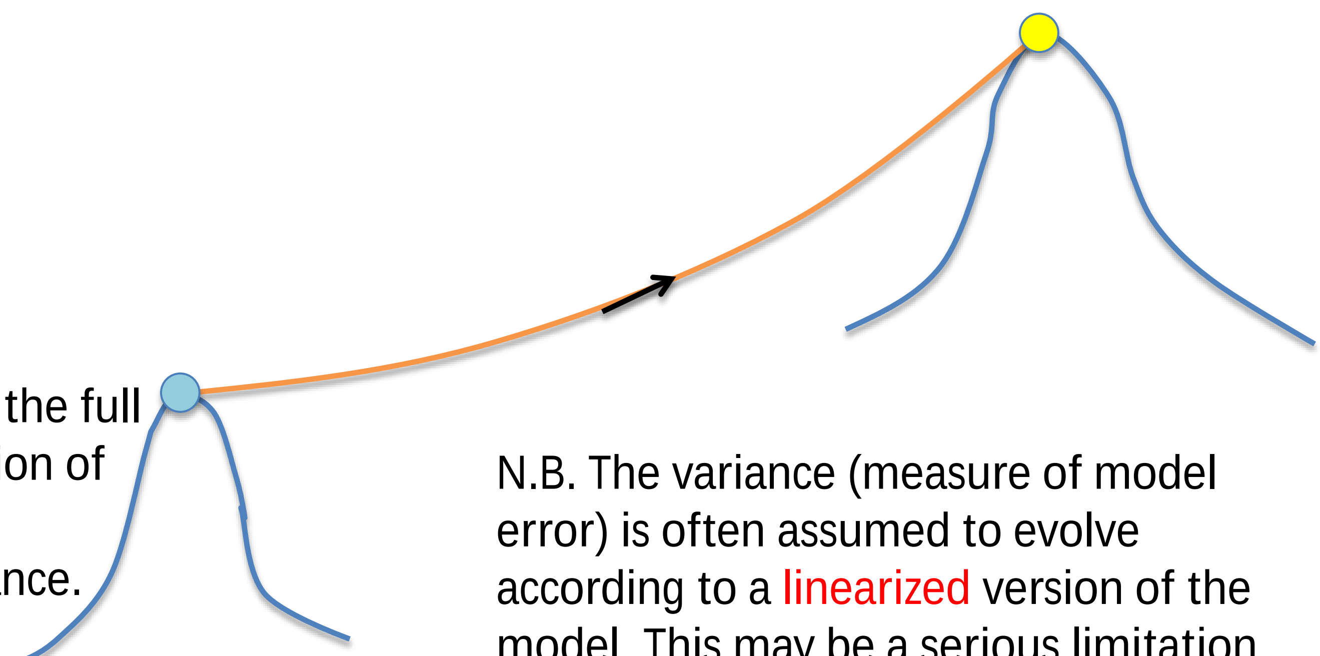


The state estimation problem



pdf evolved by nonlinear model
(Fokker-Planck equation)

Update with observations at time $k+1$ using Bayes's theorem



Too expensive to evolve the full pdf. Estimate the evolution of the low order moments: typically mean and variance. This is equivalent to assuming Gaussian statistics.

N.B. The variance (measure of model error) is often assumed to evolve according to a linearized version of the model. This may be a serious limitation for data assimilation.

State Estimation

A formula (or algorithm) to estimate the value of a state variable *x* is called an estimator. (Note: the estimator is a random variable, because it is expressed in terms of random variables, such as *y*.)

We often derive our estimator by constructing a COST FUNCTION, *J*, which measures the fit of our state variable(s) *x* to the data. Then we minimize this cost function to obtain the "optimal" *x*.

For typically used cost functions, our estimator is:

$$\hat{x} = E[x \mid y] = \int x p(x \mid y) dx$$

the mean of *x* given *y*.

For Gaussian statistics, we get our estimator $\hat{x}$ as the *x* that maximizes $p(x \mid y)$ (the mode) or which minimizes J=-ln $p(x \mid y)$.

A Simple Example

Assume we have an observation x_o of an unknown variable x.

Assume we have some prior information that the value of x is x_b.

Assume we know the error statistics of these quantities (the error variances).

$$p(x \mid x_o) \sim \underline{p(x_o \mid x)}\, p(x) = \exp\left\{ -\frac{(x_o - x)^2}{2\sigma_o^2} \right\} \exp\left\{ -\frac{(x - x_b)^2}{2\sigma_b^2} \right\}$$

$$J(x) = -\ln p(x \mid x_o) \sim \left\{ \frac{(x_o - x)^2}{2\sigma_o^2} \right\} + \left\{ \frac{(x - x_b)^2}{2\sigma_b^2} \right\}$$

Find x such that $J(x)$ is a minimum. This x is our estimate of x. Call it $\hat{x}$.

$\hat{x} \sim \frac{x_o}{\sigma_o^2} + \frac{x_b}{\sigma_b^2}$. Easy to show from form of $p(x \mid x_0)$ that $\frac{1}{\hat{\sigma}^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}$

The bigger the variance, the less weight is given to the information.

The precision of the estimate is better than those of the observation or background.

(To get an equals sign in the above, divide by the sum of the weights.)

The Observation Operator

The observations (observation vector) are in general not direct measurements of the state variables (state vector), e.g. in remote sensing from space.
In data assimilation, we need to compare the observation vector with the state vector. The observation operator allows this.
It is a mapping from state space to observation space.

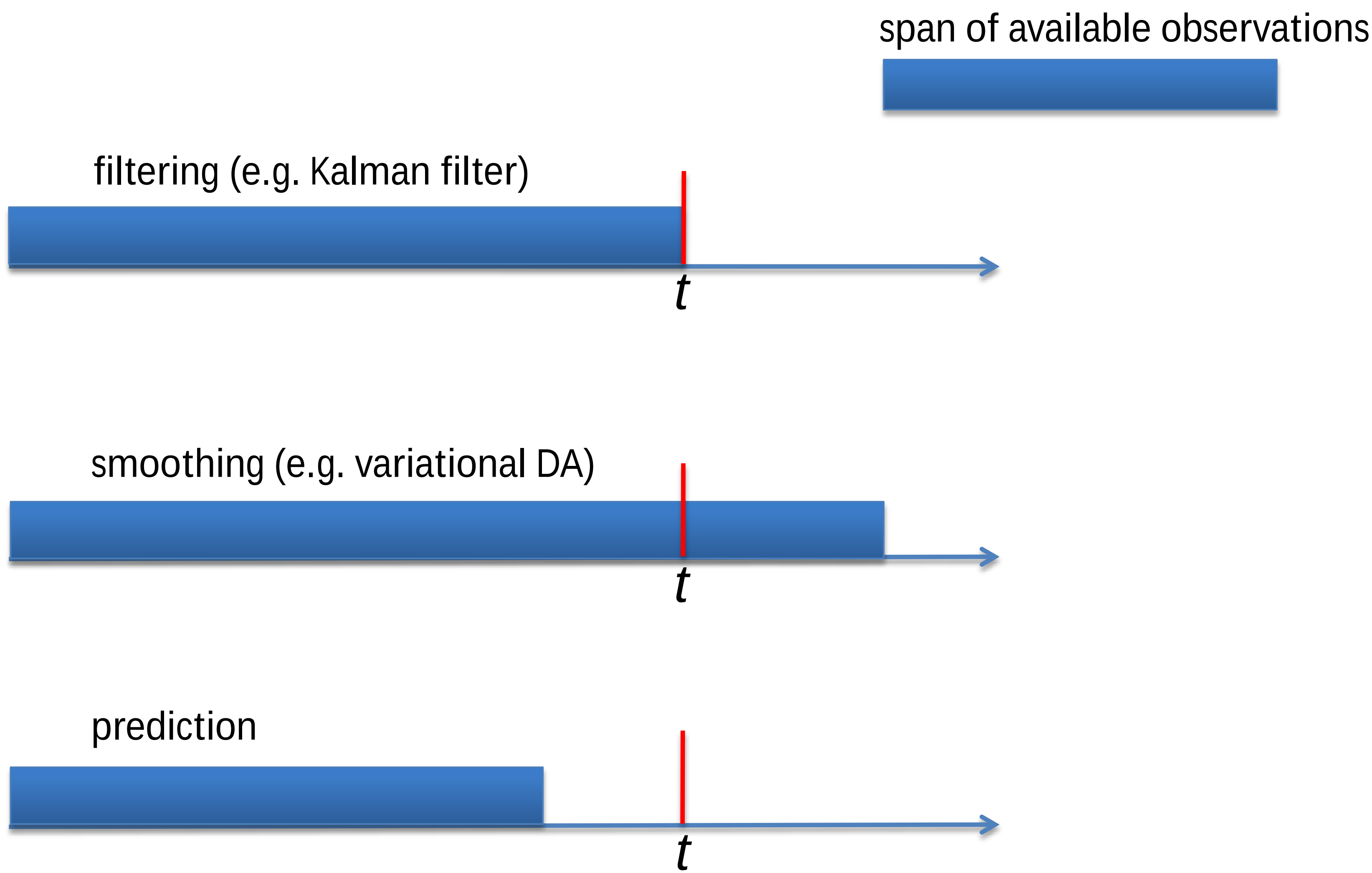
$$\mathbf{y}^{\mathrm{mod}} = h(\mathbf{x})$$

$$R_i = \int B_i(T(p))\frac{dt}{dp}$$

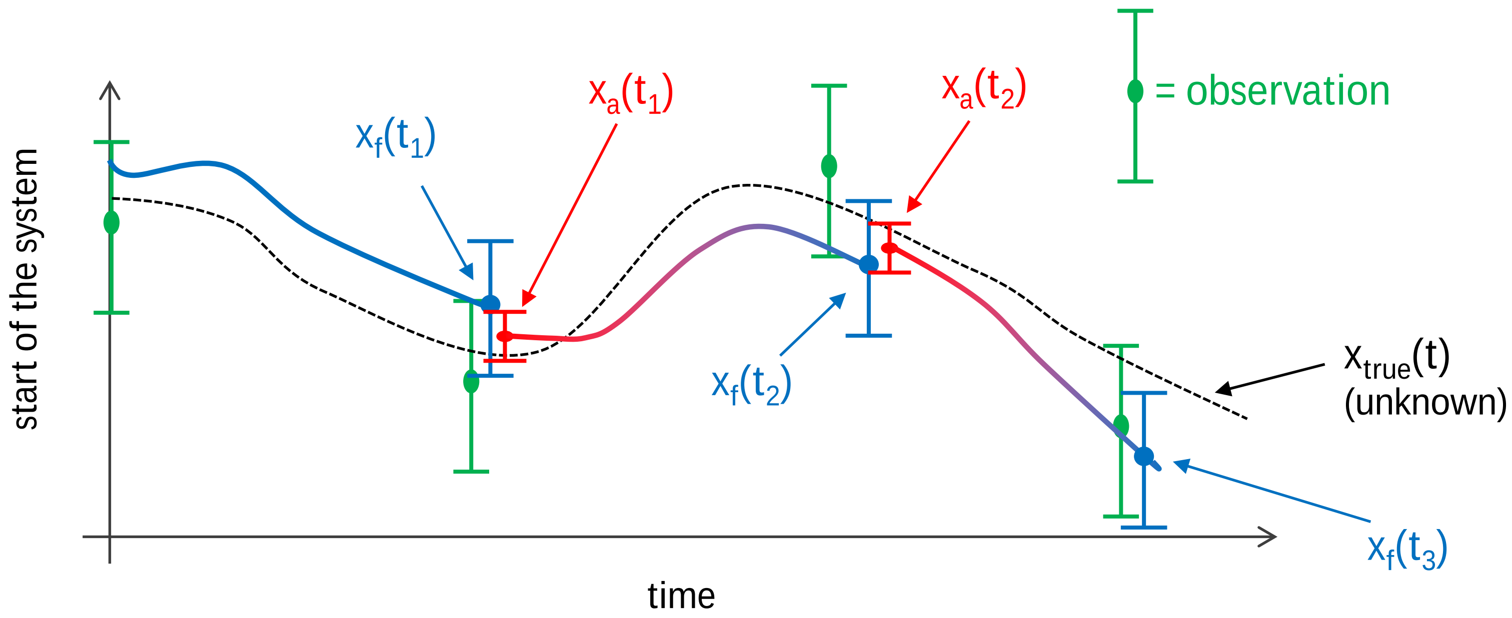
Data assimilation algorithms often use the matrix evaluated generally at a state forecast by the model (background state or first-guess state)

$$\mathbf{H} = \left.\frac{\partial h}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_B}$$

Three types of estimation problem (estimate desired at time *t*)



Sequential Data Assimilation

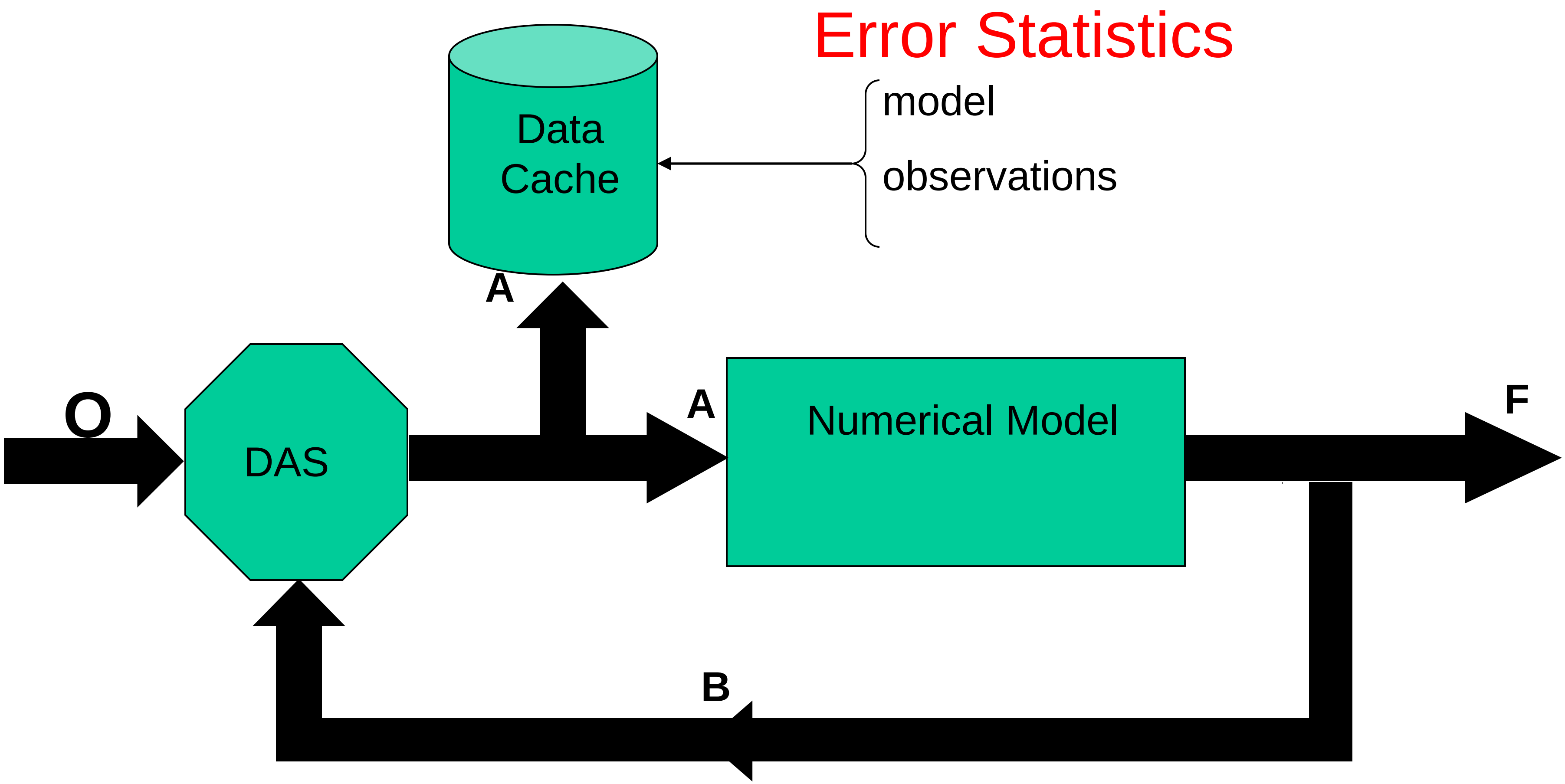


This is an example of a 'filter'

Data assimilation has:

- prediction stages (x_f = 'forecast', 'prior', 'background') (extrapolation)
- analysis stages (x_a) (interpolation)

DATA ASSIMILATION SYSTEM



3d-Variational Data Assimilation

Multivariate Case

state vector $$\mathbf{x}(t) = \begin{pmatrix} x_1 \\ x_2 \\ \bullet \\ \bullet \\ x_n \end{pmatrix}$$

observation vector $$\mathbf{y}(t) = \begin{pmatrix} y_1 \\ y_2 \\ \bullet \\ y_m \end{pmatrix}$$

Errors

The Error Covariance Matrix

$$e = \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{pmatrix} \qquad e^{\mathrm{T}} = \begin{pmatrix} e_1 & e_2 & . & . & . & e_n \end{pmatrix}$$

$$< e_i e_i > = \sigma_i^2$$

$$\mathbf{P} = < e e^{\mathrm{T}} > = \begin{pmatrix} <e_1 e_1> & <e_1 e_2> & . & . & . & <e_1 e_n> \\ <e_2 e_1> & <e_2 e_2> & . & . & . & <e_2 e_n> \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ <e_n e_1> & <e_n e_2> & . & . & . & <e_n e_n> \end{pmatrix}$$

Background Errors

- They are the estimation errors of the background state (a model forecast):

$$e_b = \mathbf{x}_b - \mathbf{x}$$

- average (bias) $< e_b >$
- covariance

$$\mathbf{B} = < (e_b - < e_b >)(e_b - < e_b >)^T >$$

Background error "in observation space"

If $\mathbf{y}^{\mathrm{mod}} = \mathbf{H}\mathbf{x}_b$ where $\mathbf{H}$ is a matrix, then the error covariance for $\mathbf{y}^{\mathrm{mod}}$ is given by:

$$\mathbf{C}_{\mathbf{y}^{\mathbf{mod}}} = \mathbf{HBH}^{\mathrm{T}}$$

Observation Errors

- They contain errors in the observation process (instrumental error), errors in the design of H, and "representativeness errors", i.e. discretizaton errors that prevent $\mathbf{x}$ from being a perfect representation of the true state.

$$e_o = \mathbf{y} - H(\mathbf{x})$$

$$\mathbf{R} = \langle (e_o - \langle e_o \rangle)(e_o - \langle e_o \rangle)^T \rangle$$

The Bayesian view of data assimilation

<u>Bayes' Theorem</u>

$$\left.\begin{aligned} P(\vec{y},\vec{x}) &= P(\vec{x}\,|\,\vec{y})P(\vec{y}) \\ P(\vec{x},\vec{y}) &= P(\vec{y}\,|\,\vec{x})P(\vec{x}) \end{aligned}\right\} \quad P(\vec{x}\,|\,\vec{y}) = \frac{P(\vec{y}\,|\,\vec{x})P(\vec{x})}{P(\vec{y})}$$

$$\propto P(\vec{x})P(\vec{y}\,|\,\vec{x})$$

Rev. Thomas Bayes 1702-1761

$$P(\vec{x}\,|\,\vec{y}) \propto \exp\left(-\frac{1}{2}(\vec{x}-\vec{x}_B)^T \mathbf{B}^{-1}(\vec{x}-\vec{x}_B)\right)\exp\left(-\frac{1}{2}(\vec{h}[\vec{x}]-\vec{y})^T \mathbf{R}^{-1}(\vec{h}[\vec{x}]-\vec{y})\right)$$

$$\propto \exp-\left(\frac{1}{2}(\vec{x}-\vec{x}_B)^T \mathbf{B}^{-1}(\vec{x}-\vec{x}_B) + \frac{1}{2}(\vec{h}[\vec{x}]-\vec{y})^T \mathbf{R}^{-1}(\vec{h}[\vec{x}]-\vec{y})\right)$$

Maximum likelihood $\Rightarrow$ Minimum penalty, J

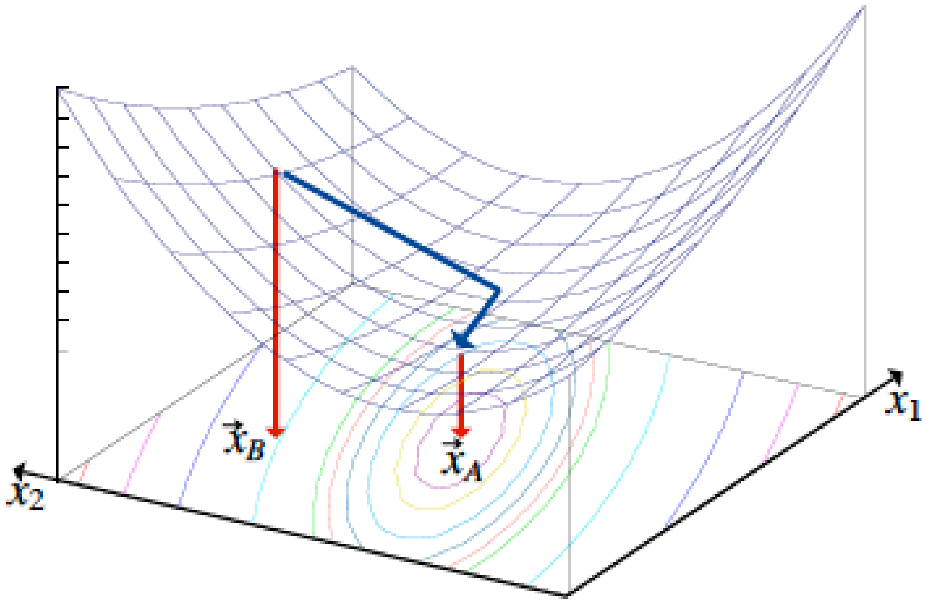
$$J[\vec{x}] = \frac{1}{2}(\vec{x}-\vec{x}_B)^T \mathbf{B}^{-1}(\vec{x}-\vec{x}_B) + \frac{1}{2}(\vec{h}[\vec{x}]-\vec{y})^T \mathbf{R}^{-1}(\vec{h}[\vec{x}]-\vec{y})$$

$$\vec{x}_A = \vec{x}\big|_{\min J} = \text{"analysis"}$$

Minimising the cost function

The problem reduces to a (badly conditioned) optimisation problem in 10^7-dimensional phase space.

$$J[\vec{x}] = \frac{1}{2}(\vec{x} - \vec{x}_B)^T \mathbf{B}^{-1} (\vec{x} - \vec{x}_B) + \frac{1}{2}(\vec{h}[\vec{x}] - \vec{y})^T \mathbf{R}^{-1} (\vec{h}[\vec{x}] - \vec{y})$$



- Descent algorithms minimize J iteratively.
- They need the local gradient, $\nabla_{\vec{x}} J$ of the cost function at each iteration.
- The adjoint method is used to compute the adjoint.
- The curvature^{-1} (a.k.a. inverse Hessian, $(\nabla_{\vec{x}}^2 J)^{-1}$) at $\vec{x}_A$ indicates the error statistics of the analysis.
  - A very badly conditioned problem.

Remarks on 3d-VAR

- Can add constraints to the cost function, e.g. to help maintain “balance”
- Can work with non-linear observation operator *H.*
- Can assimilate radiances directly (simpler observational errors).
- Can perform global analysis instead of OI approach of radius of influence.

Optimal Interpolation (the BLUE)

- BLUE = Best linear unbiased estimate
- Algorithm derived as a special case of 3D-var.

Algebraic minimization of the cost function

Under simplified conditions the cost function can be minimized algebraically.

Assume that the linearization of the forward model is reasonable

$$\vec{h}[\vec{x}] \approx \vec{h}[\vec{x}_B] + \mathbf{H}(\vec{x} - \vec{x}_B)$$

$$J[\vec{x}] = \frac{1}{2}(\vec{x} - \vec{x}_B)^T \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) + \frac{1}{2}(\mathbf{H}(\vec{x} - \vec{x}_B) - (y - \vec{h}[\vec{x}_B]\vec{)})^T \mathbf{R}^{-1}(\mathbf{H}(\vec{x} - \vec{x}_B) - (y - \vec{h}[\vec{x}_B]\vec{)})$$

1. Calculate the gradient vector

$$\nabla_{\vec{x}} J = \begin{pmatrix} \partial J / \partial x_1 \\ \partial J / \partial x_2 \\ \\ \partial J / \partial x_N \end{pmatrix} = \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) + \mathbf{H}^T \mathbf{R}^{-1}(\vec{h}[\vec{x}] - \vec{y})$$

2. The special $\vec{x}$ that has zero gradient minimizes J (this cost function is quadratic and convex)

$$\nabla_{\vec{x}} J\big|_{x_A} = 0$$

$$\vec{x}_A = \vec{x}_B + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}(\vec{y} - \vec{h}[\vec{x}_B])$$

$$= \vec{x}_B + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\vec{y} - \vec{h}[\vec{x}_B])$$

This is the OI formula with the BLUE!

BLUE Estimator (recursive)

- The BLUE estimator or "analysis" is given by:

$$\mathbf{x}_{\mathrm{a}} = \mathbf{x}_{\mathrm{b}} + \mathbf{K}(\mathbf{y} - h(\mathbf{x}_{\mathrm{b}}))$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

- The $\mathbf{B}$ matrix plays a key role in determining the structure of the analysed fields.
- Matrix inverses expensive to compute so reduce dimension by "local analysis"
- We can derive an explicit expression for the analysis error covariance matrix:

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}$$

Assumptions Used in *BLUE*

- Linearized observation operator:

$$h(\mathbf{x}) - h(\mathbf{x}_\mathrm{b}) = \mathbf{H}(\mathbf{x} - \mathbf{x}_\mathrm{b})$$

- Errors are unbiased:

$$< \mathbf{x}_\mathrm{b} - \mathbf{x} > = < \mathbf{y} - h(\mathbf{x}) > = 0$$

- Errors are uncorrelated:

$$< (\mathbf{x}_\mathrm{b} - \mathbf{x})(\mathbf{y} - h(\mathbf{x}))^\mathrm{T} > = 0$$

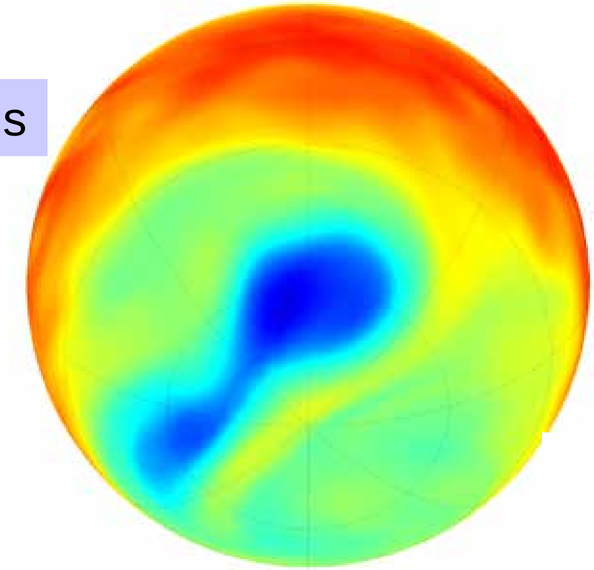
Innovations and Residuals

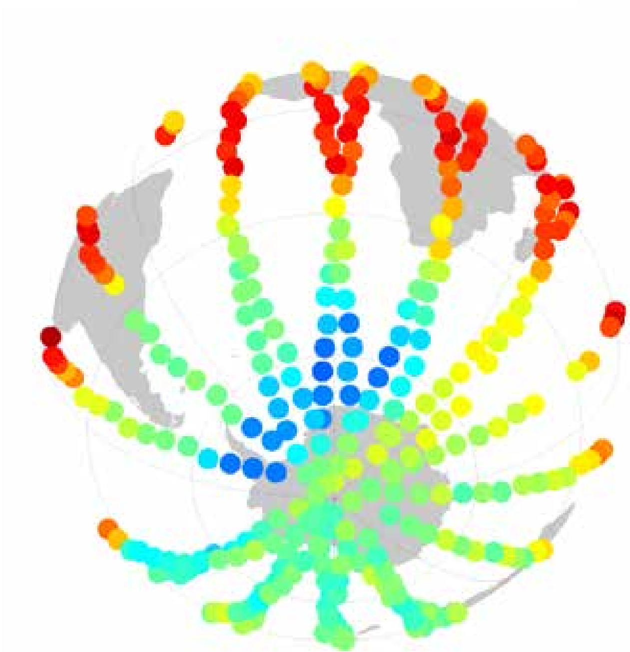
- Key to data assimilation is the use of differences between observations and the state vector of the system
- We call $\mathbf{y} - h(\mathbf{x}_b)$ the innovation
- We call $\mathbf{y} - h(\mathbf{x}_a)$ the analysis residual

Give important information

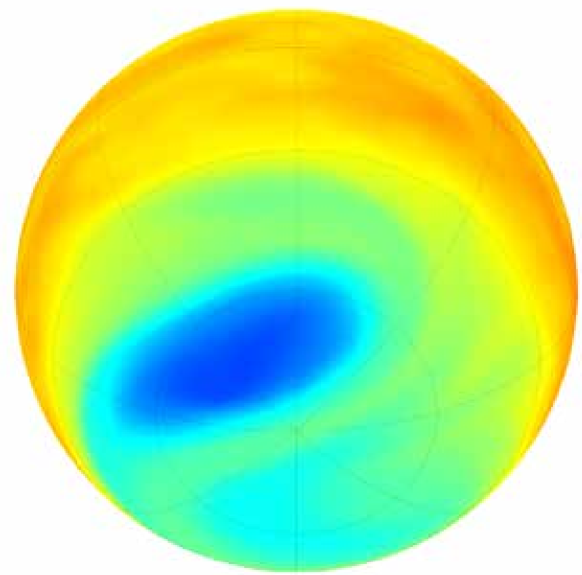
Ozone at 10hPa, 12Z 23rd Sept 2002

Analysis





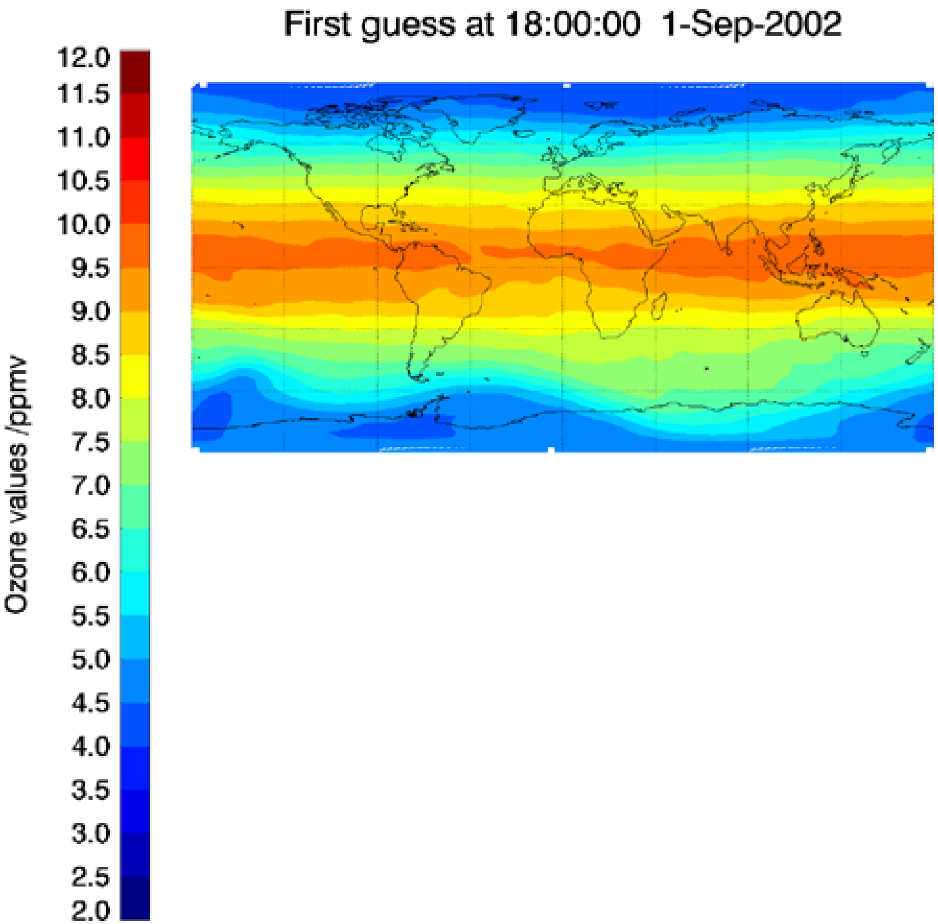
MIPAS observations



6 day model forecast

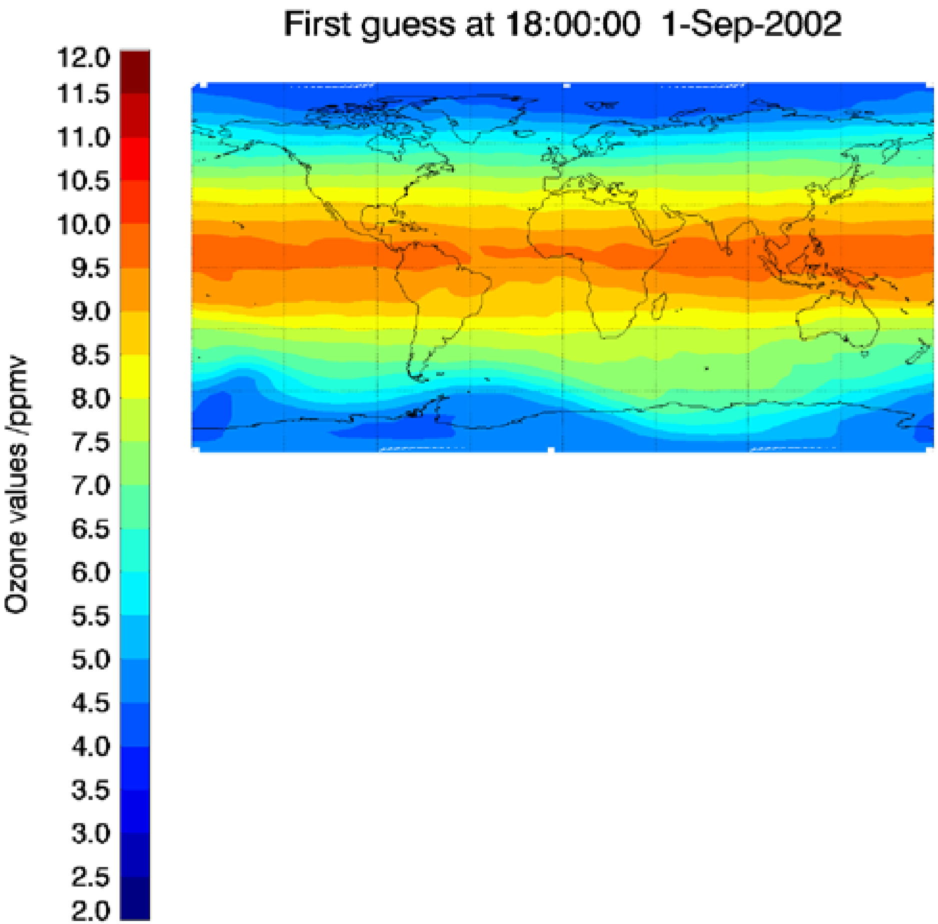
3D variational data assimilation - ozone at 10hPa

$\mathbf{x}_b$

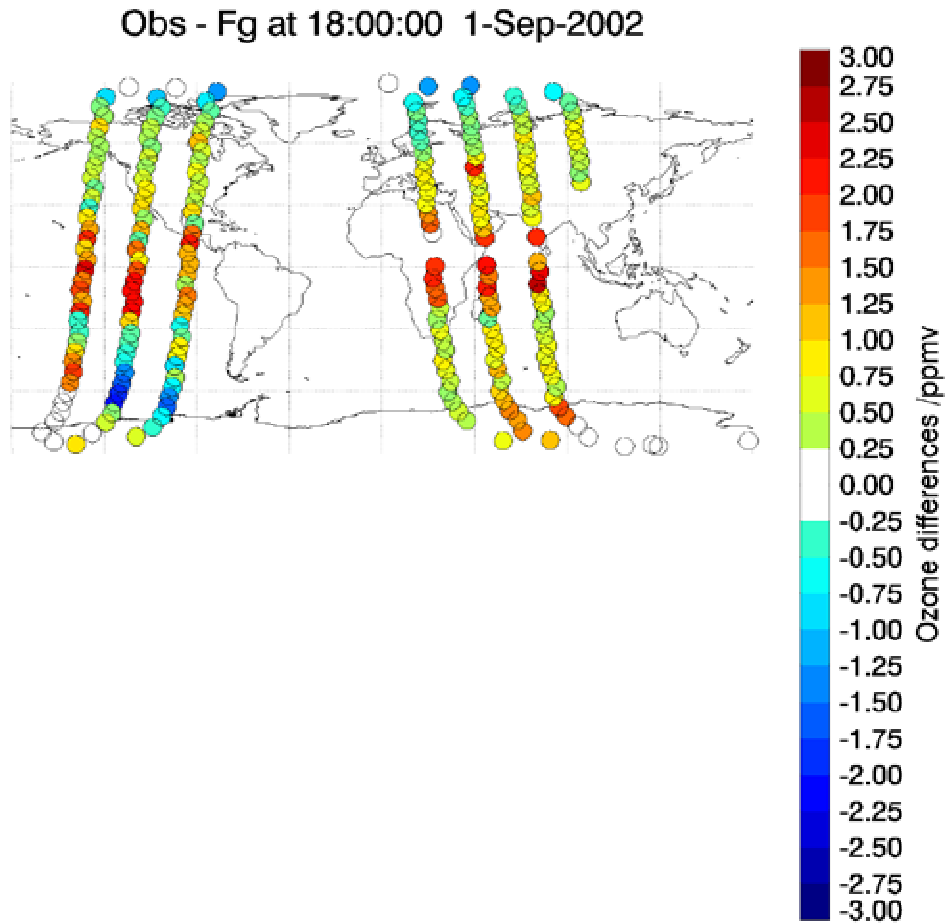


3D variational data assimilation - ozone at 10hPa

$\mathbf{x}_b$

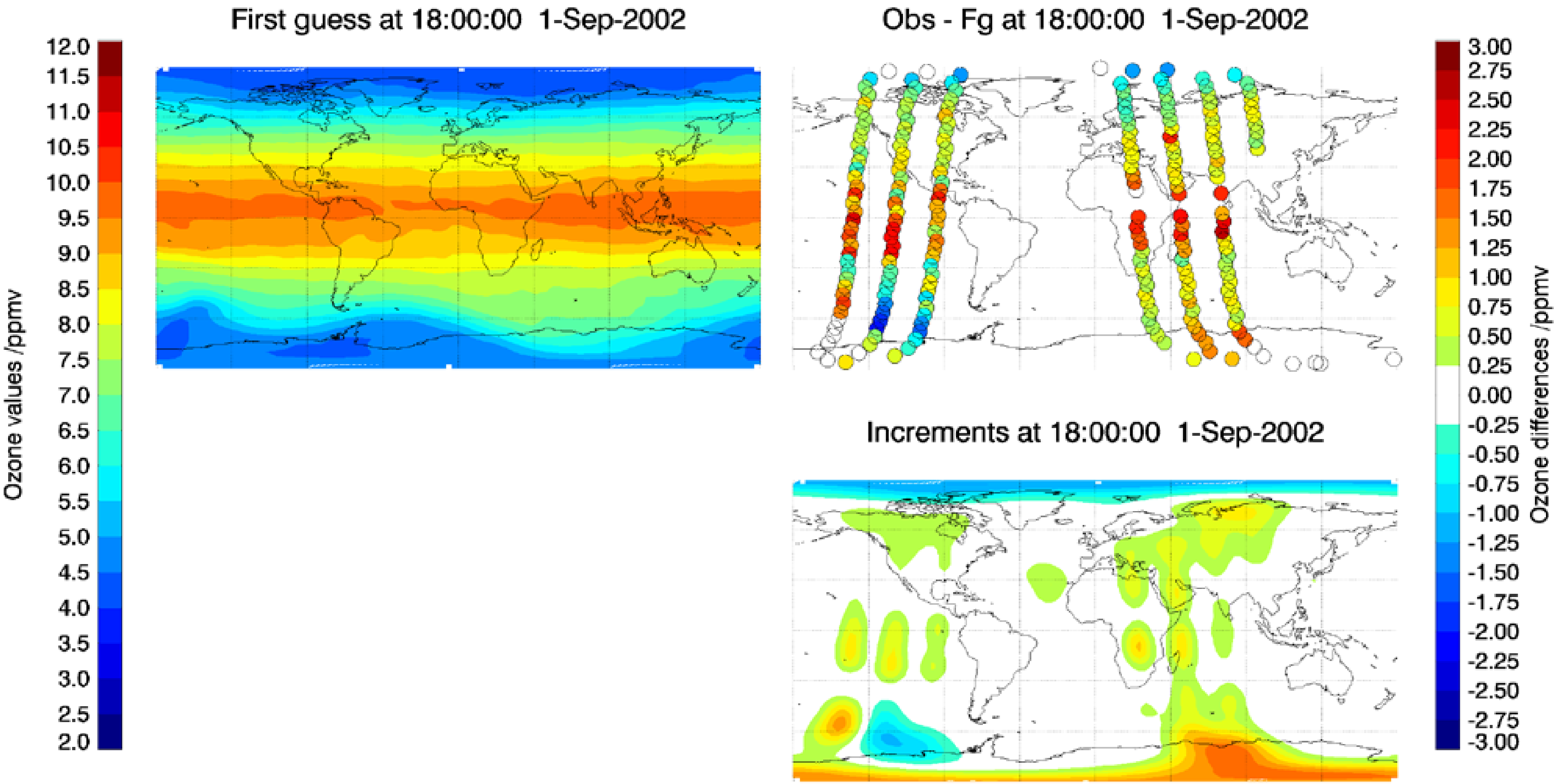


$\mathbf{y} - h(\mathbf{x}_b)$



3D variational data assimilation - ozone at 10hPa

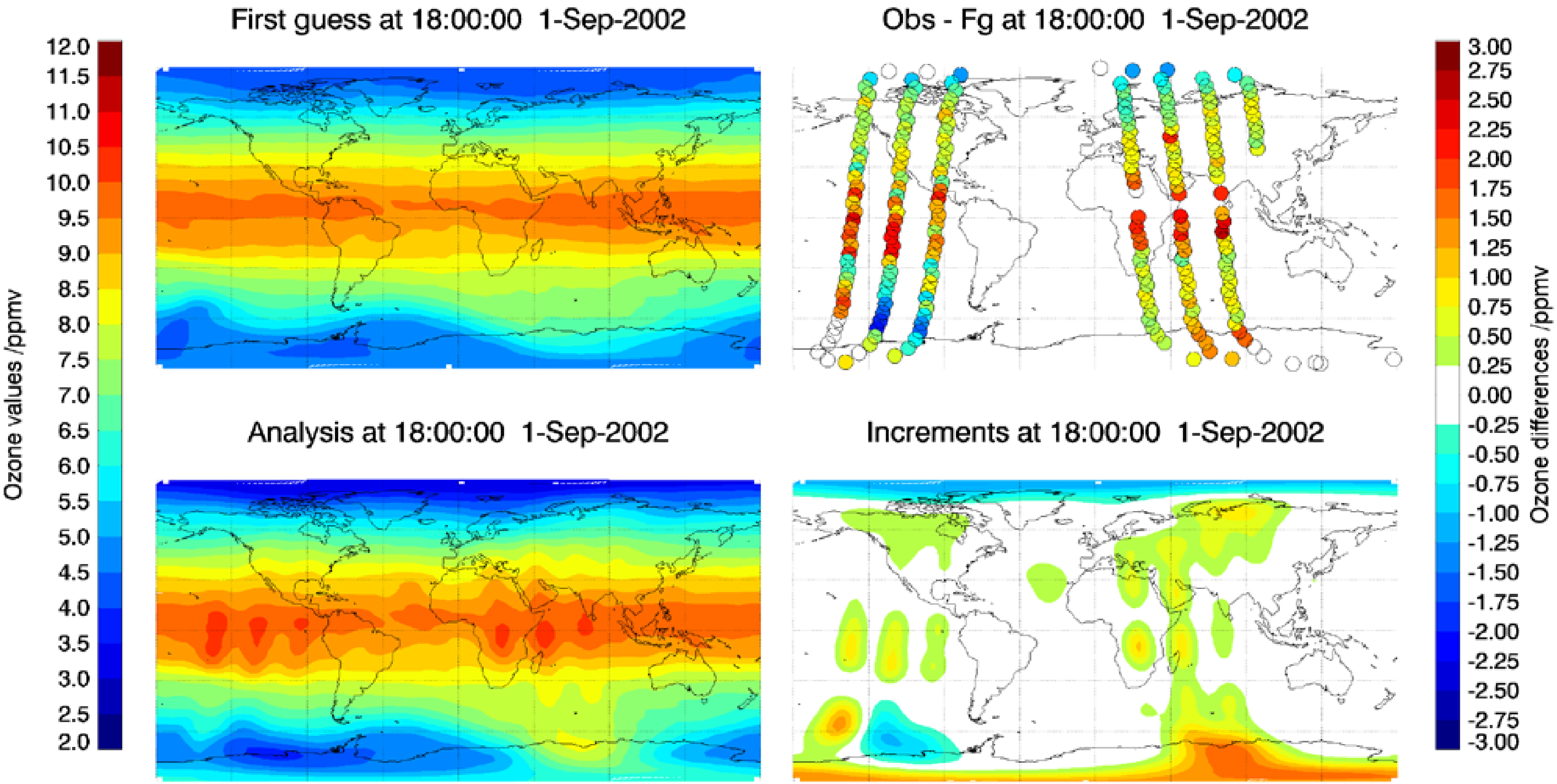
$\mathbf{x}_b$ $\qquad\qquad$ $\mathbf{y} - h(\mathbf{x}_b)$



$\mathbf{K}(\mathbf{y} - h(\mathbf{x}_b))$

The data assimilation cycle: ozone at 10hPa

$$\mathbf{x}_b \qquad\qquad \mathbf{y} - h(\mathbf{x}_b)$$

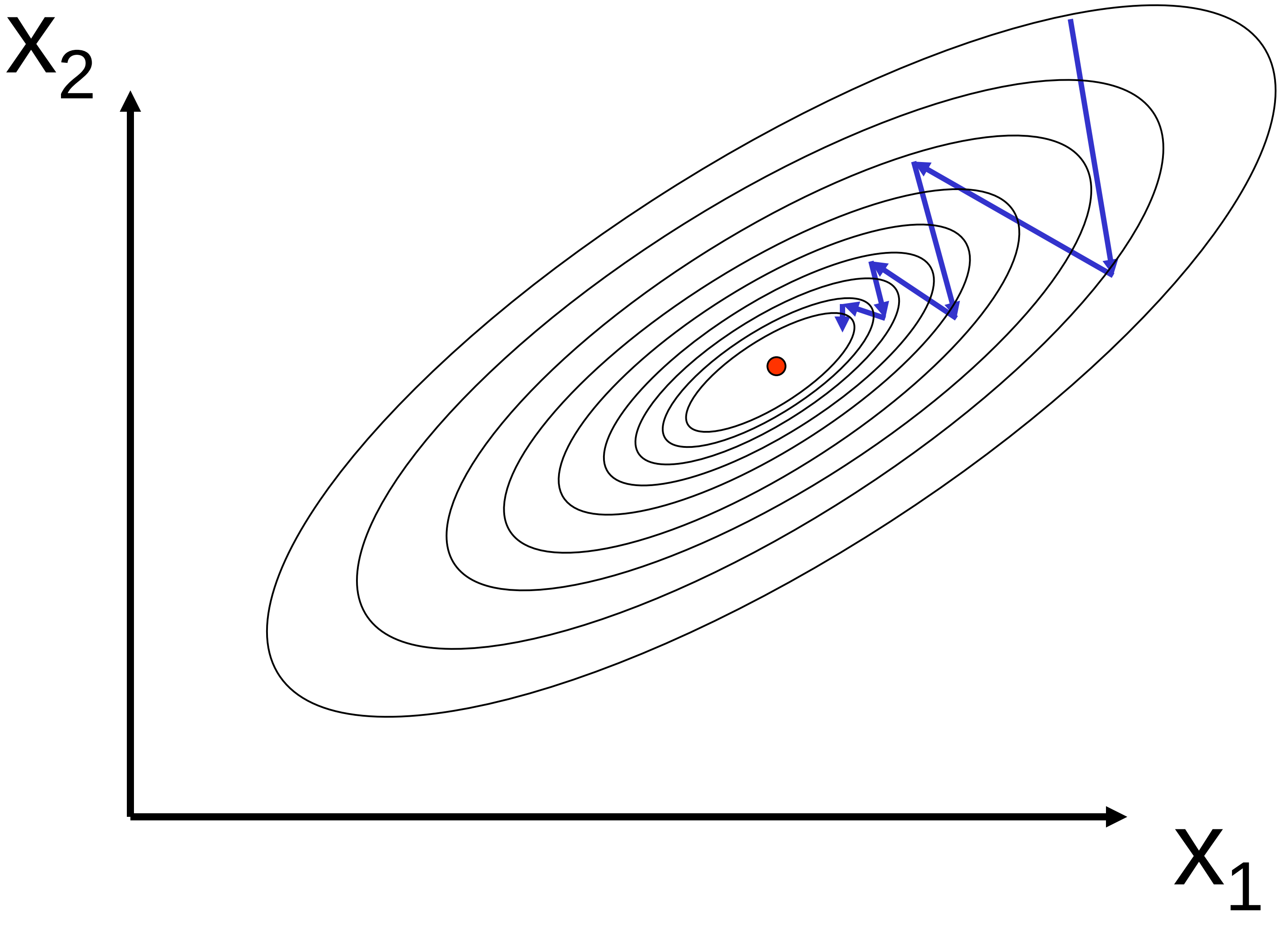


$$\mathbf{x}_b + \mathbf{K}(\mathbf{y} - h(\mathbf{x}_b)) \qquad\qquad \mathbf{K}(\mathbf{y} - h(\mathbf{x}_b))$$

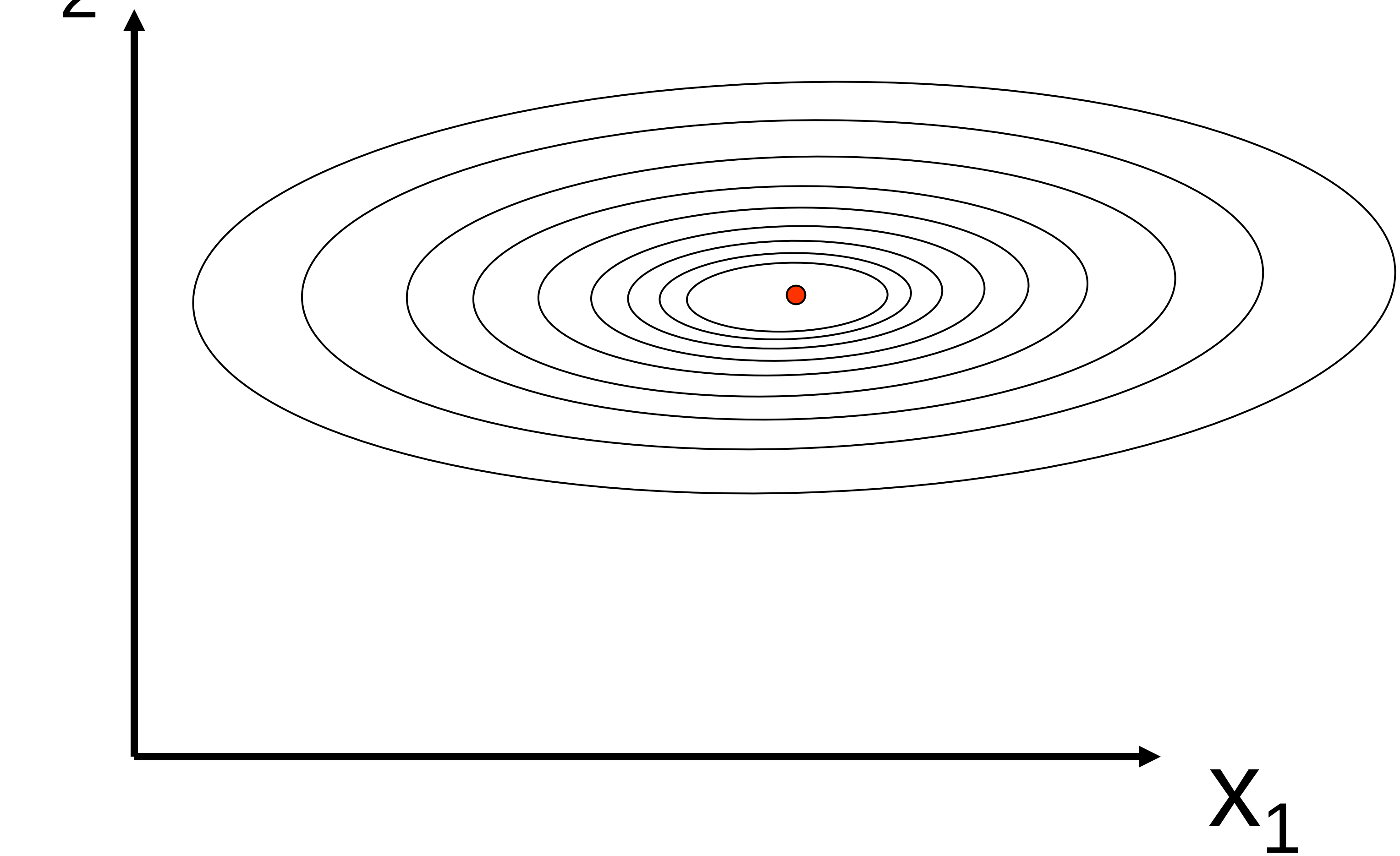
Choice of State Variables and Preconditioning

- Free to choose which variables to use to define state vector, $x(t)$
- We'd like to make B diagonal
  - may not know *covariances* very well
  - want to make the minimization of J more efficient by "preconditioning": transforming variables to make surfaces of constant J nearly spherical in state space

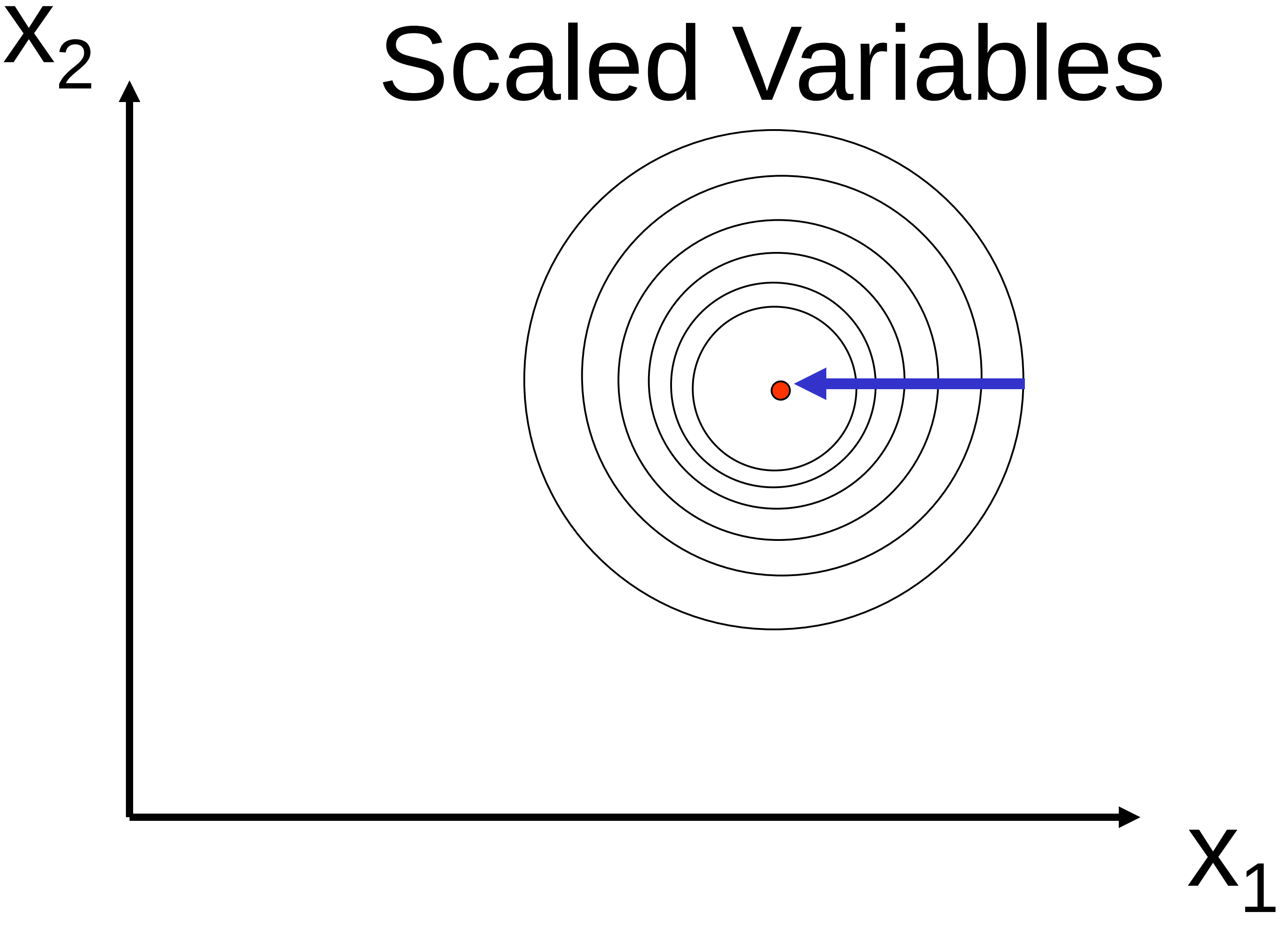
Cost Function for Correlated Errors



Cost Function for Uncorrelated Errors



Cost Function for Uncorrelated Errors Scaled Variables



END