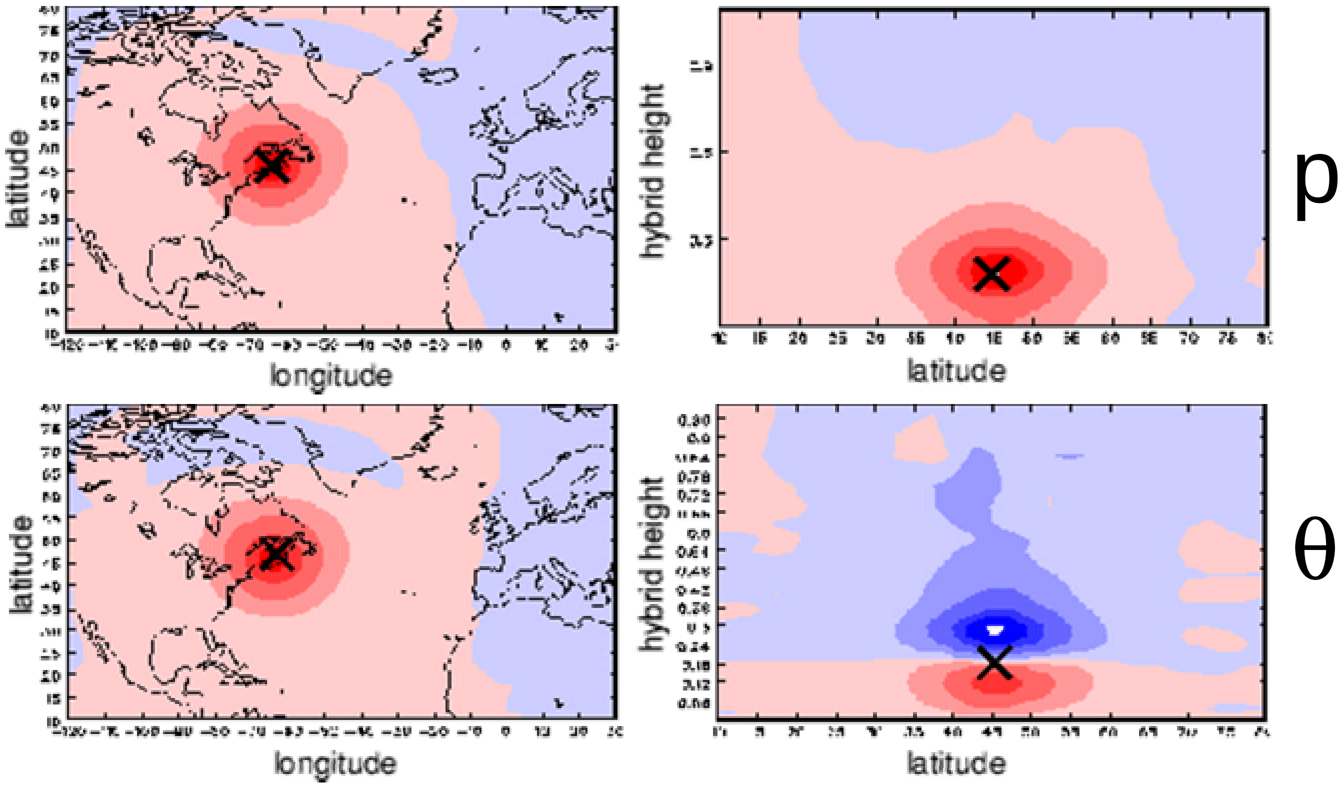
Data Assimilation

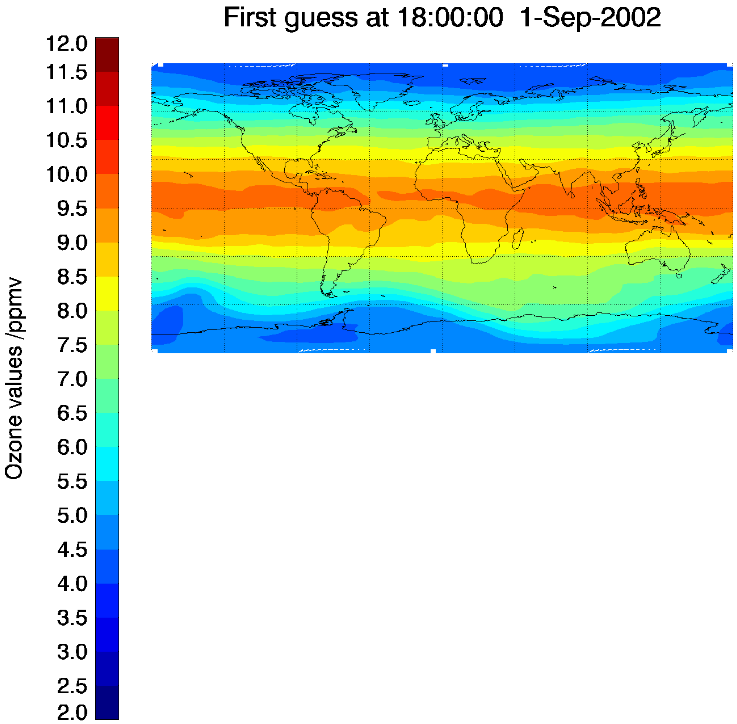
Alan O'Neill
Data Assimilation Research Centre
University of Reading

Spreading of Information from Single Pressure Obs.



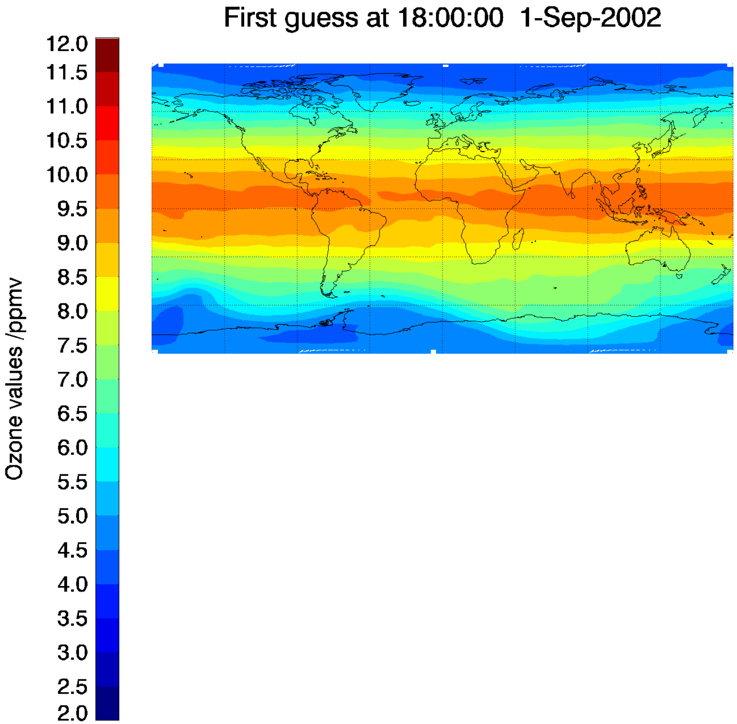
3D variational data assimilation - ozone at 10hPa

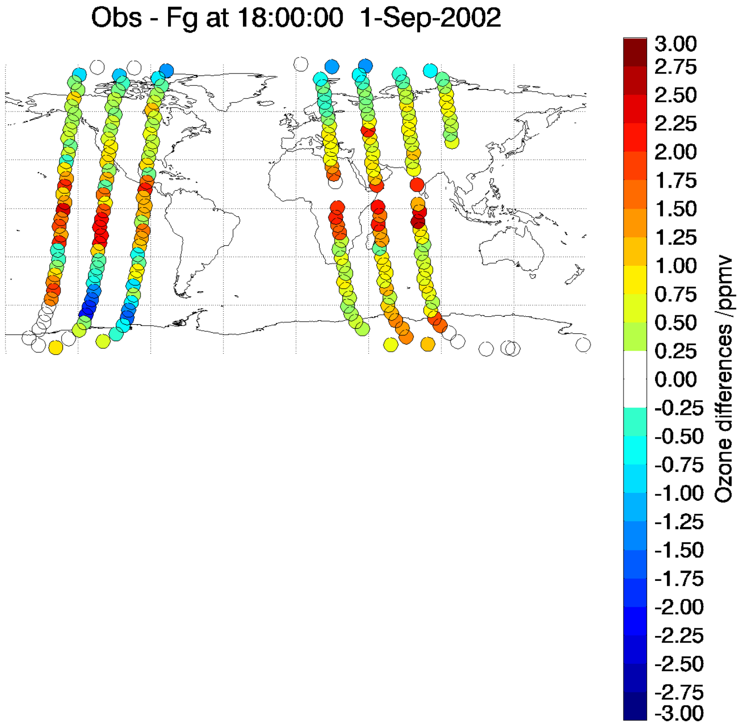
$\mathbf{x}_b$



3D variational data assimilation - ozone at 10hPa

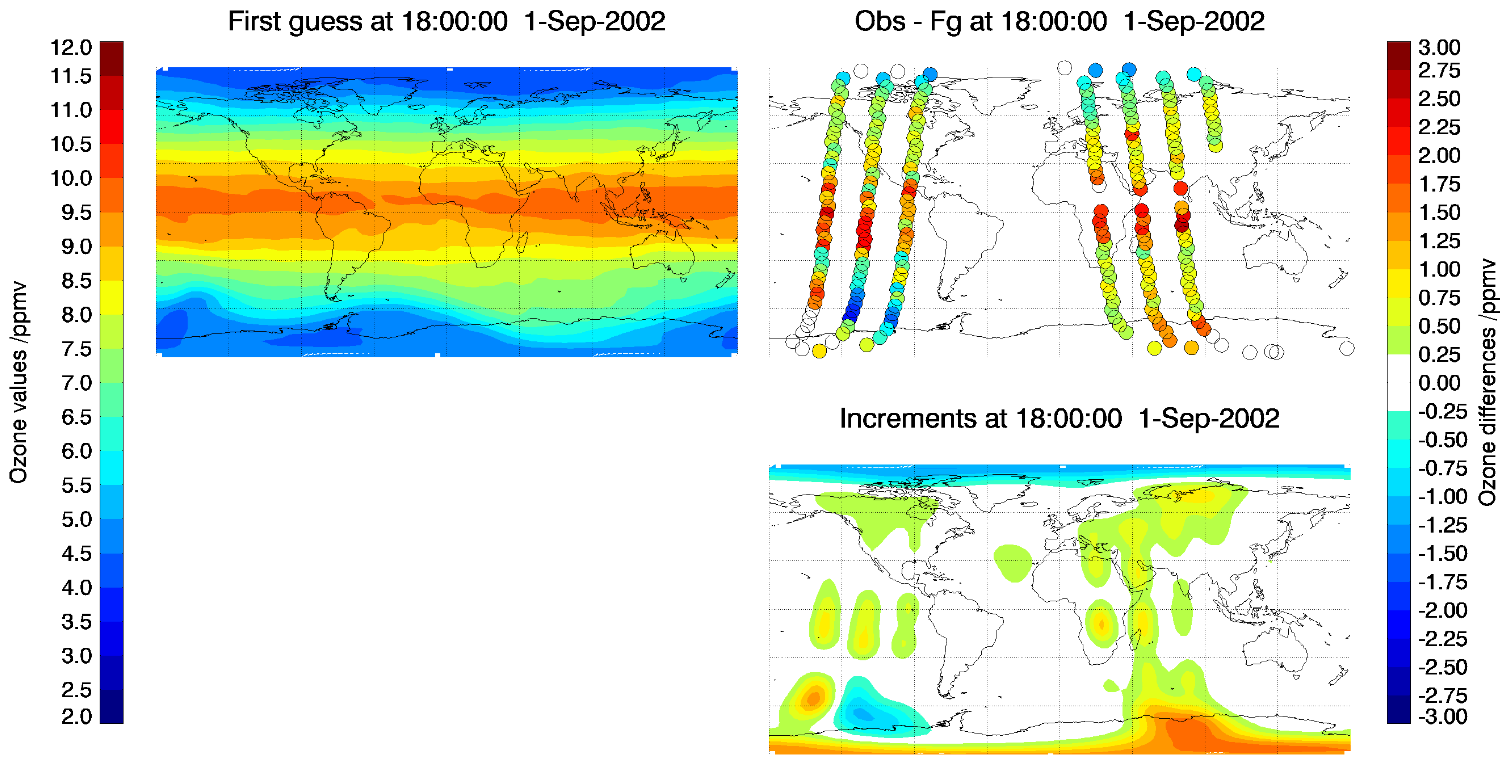
$\mathbf{x}_b$ $\qquad\qquad$ $\mathbf{y} - h(\mathbf{x}_b)$





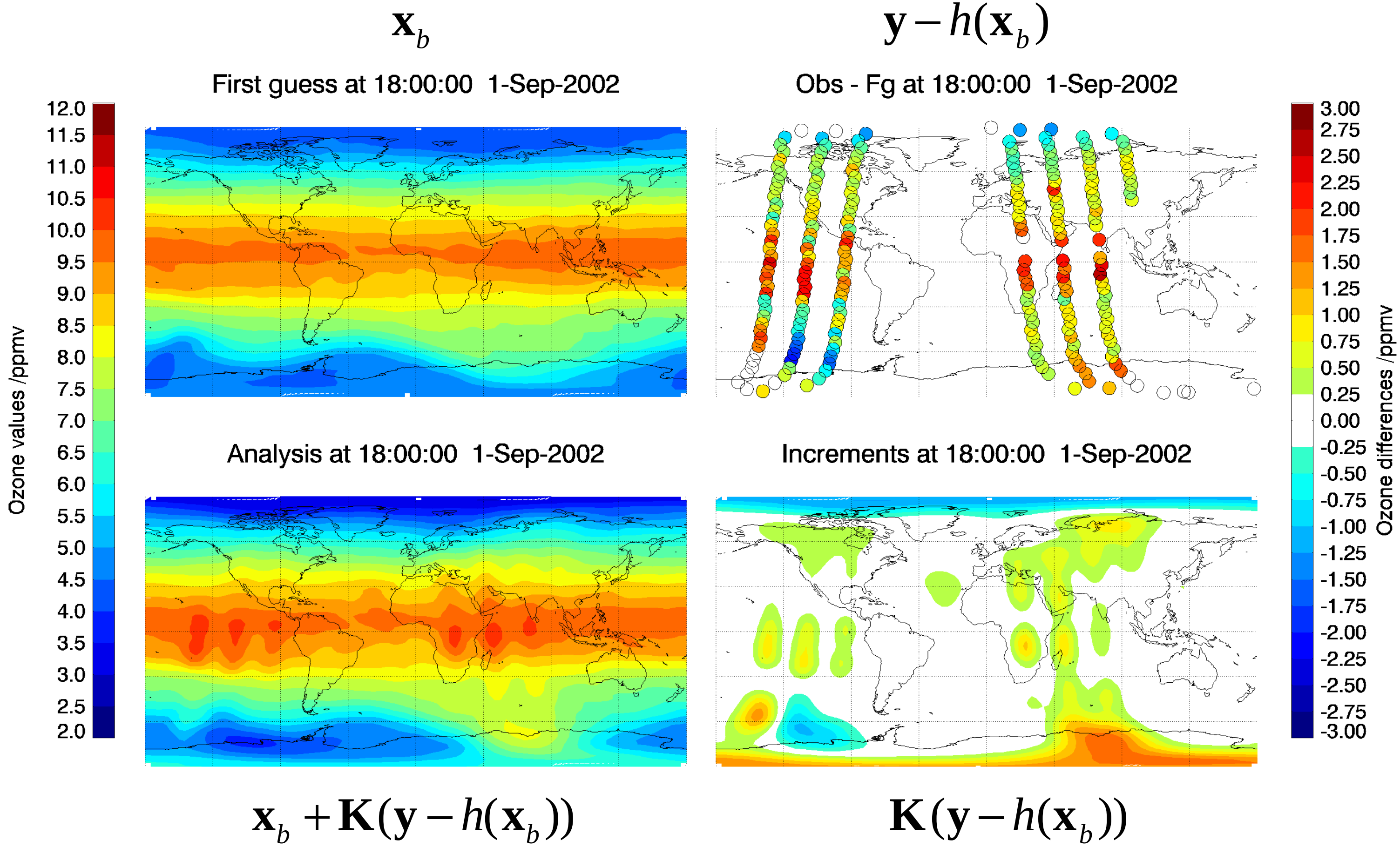
3D variational data assimilation - ozone at 10hPa

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



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

The data assimilation cycle: ozone at 10hPa



Estimating Error Statistics

- Error variances reflect our uncertainty in the observations or background.
- Often assume they are stationary in time and uniform over a region of space.
- Can estimate by *observational method* or as *forecast differences* (NMC method).
- More advanced, flow dependent errors estimated by *Kalman filter.*

Estimating Covariance Matrix for Observations, O

- O usually quite simple:
  - diagonal or
  - for nadir-sounding satellites, non-zero values between points in vertical only
- Calibration against independent measurements

Estimating the Error Covariance Matrix B

- Model B with simple functions based on comparisons of forecasts with observations:

$$B_{ij} \propto s_i s_j \exp(-d_{ij}/L) \qquad \text{horiz. fn x vert. fn}$$

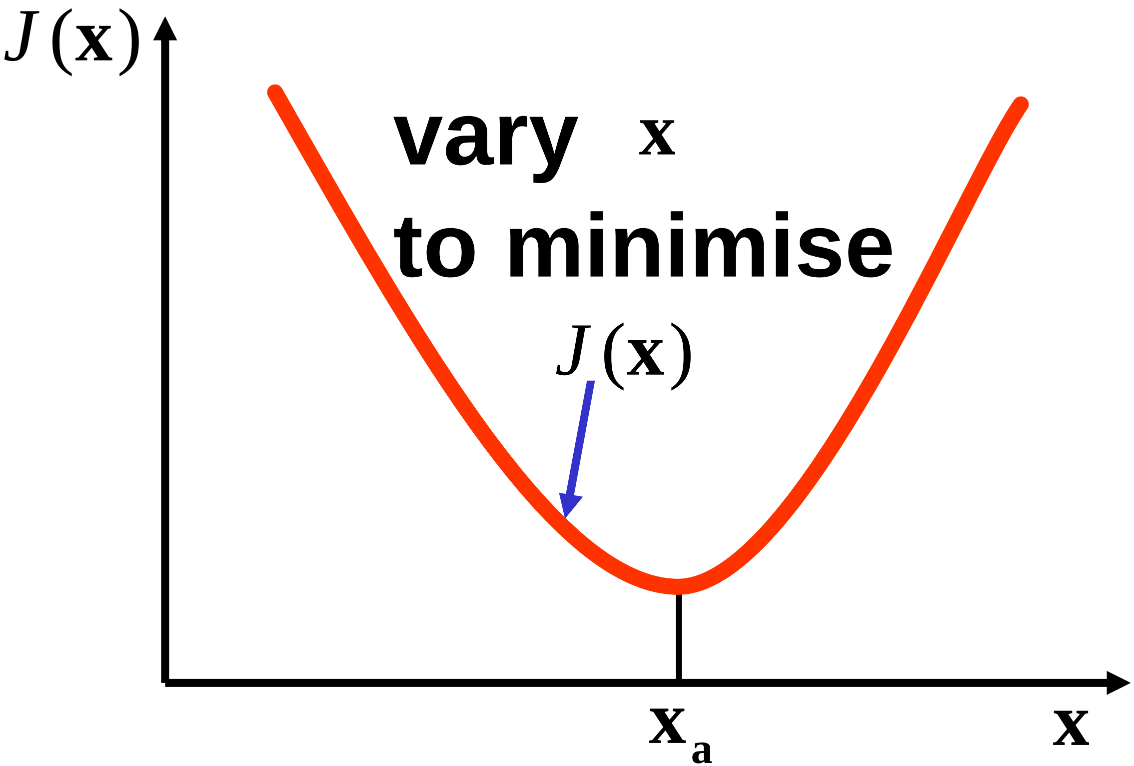
- Error growth in short-range forecasts "verifying" at the same time (NMC method)

$$\mathbf{B} \approx \langle [\mathbf{x}_f(48h) - \mathbf{x}_f(24h)][\mathbf{x}_f(48h) - \mathbf{x}_f(24h)]^{\mathrm{T}} \rangle$$

state vector at time *t* from forecast 48h or 24 h earlier

3d-Variational Data Assimilation

Variational Data Assimilation



Variational Data Assimilation

$$J(\mathbf{x}) =$$

$$(\mathbf{x}-\mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) +$$

$$(\mathbf{y}-H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))$$

nonlinear operator

assimilate y directly

global analysis

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.

Maximum Probability or Likelihood

- For Gaussian errors the background, observation and analysis pdfs are:

$$P_{\mathrm{b}}(\mathbf{x}) = \mathrm{b}\exp[(\mathbf{x}-\mathbf{x}_{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_{\mathrm{b}})/2]$$

$$P_{\mathrm{o}}(\mathbf{x}) = \mathrm{o}\exp[(\mathbf{y}-H(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))/2]$$

$$P_{\mathrm{a}}(\mathbf{x}) = \mathrm{a}\,P_{\mathrm{b}}(\mathbf{x})P_{\mathrm{o}}(\mathbf{x})$$

where b, o, and a are normalizing factors.

- Maximum probability estimate minimizes

$$J(\mathbf{x}) = -\ln P_{\mathrm{a}}(\mathbf{x})$$

Comments

- Biases occur in background and observations. Remove them if known, otherwise analysis is sub-optimal. Monitor (O-B), but is the bias in the model or in observations?
- B and O errors usually uncorrelated, but could be correlations in satellite retrievals.
- Error in the linearization of *H* should be much smaller than observational errors for all values of $\mathbf{x} - \mathbf{x}_\mathrm{b}$ met in the analysis procedure.

Effect of Observed Variables on Unobserved Variables

- Implicitly through the governing equations of the (forecast) model.
- Explicitly through the off-diagonal terms in B:

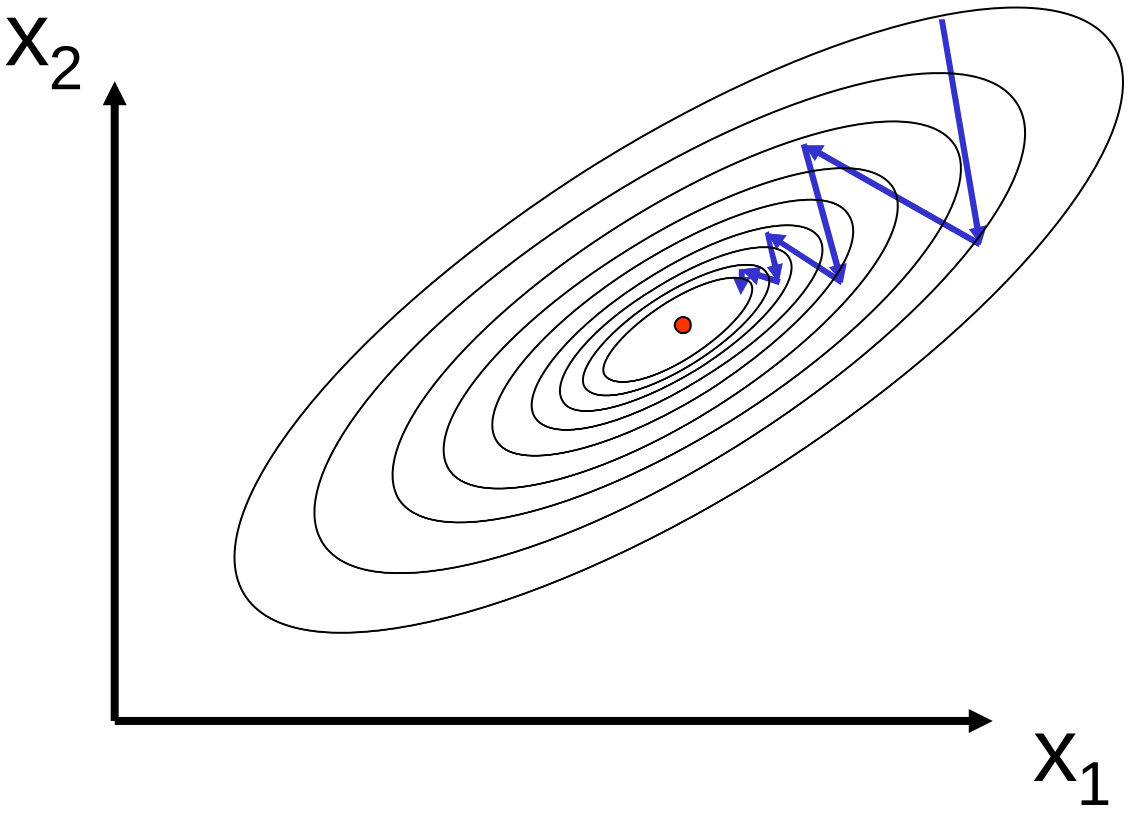
$$\begin{pmatrix} a & c \\ c & b \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}\left(y_1 - \mathbf{H}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} a & c \\ c & b \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}(y_1 - x_1) = \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix}$$

assume that y_1 is a measurement of x_1, but x_2 not measured

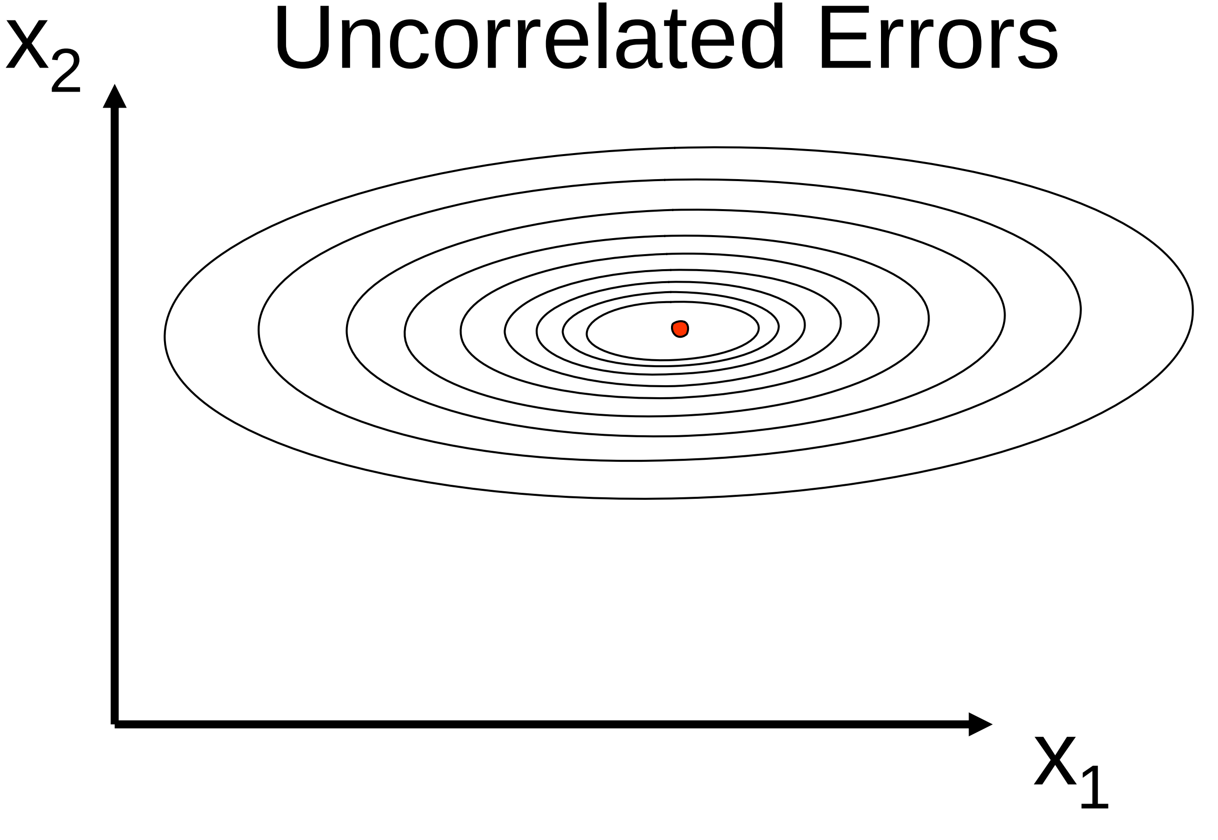
Choice of State Variables and Preconditioning

- 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

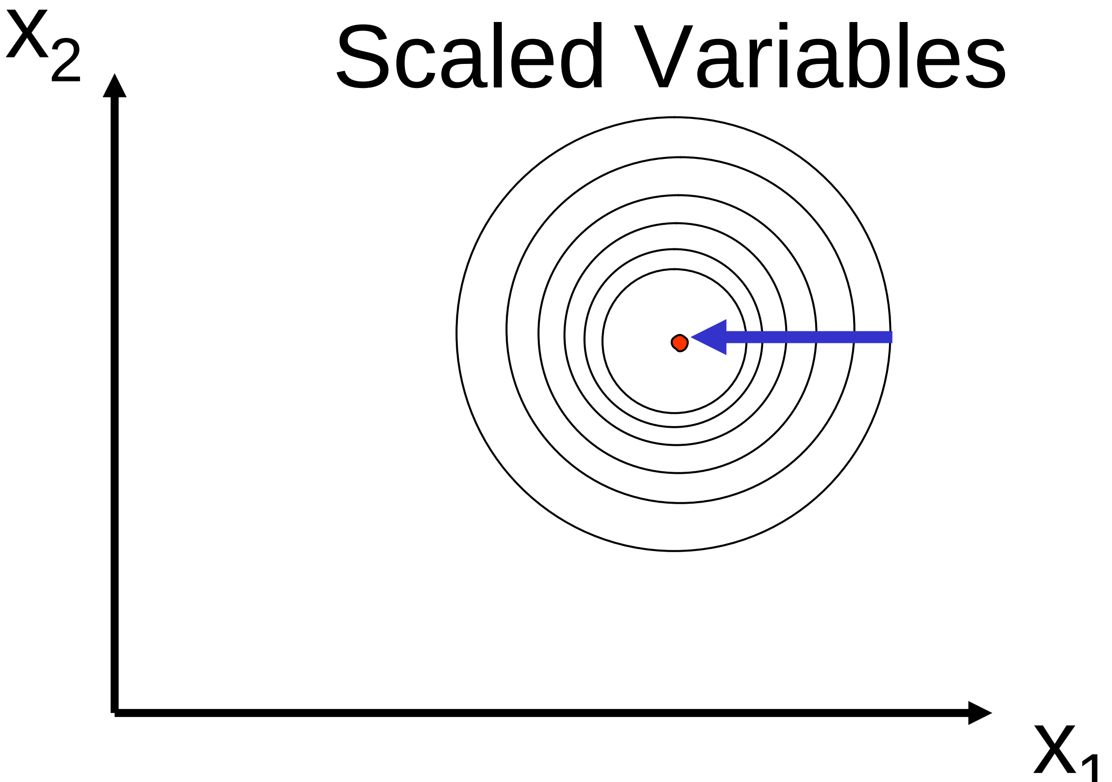
Cost Function for Correlated Errors



Cost Function for Uncorrelated Errors



Cost Function for Uncorrelated Errors Scaled Variables



The Kalman Filter

Kalman Filter

(*expensive*)

Use model equations to propagate B forward in time.

B $\longrightarrow$ B(t)

Analysis step as in OI

Evolution of Covariance Matrices

$$\mathbf{x}_b{}^{n+1} = M(\mathbf{x}_a{}^n) = M(\mathbf{x}_t{}^n + e_a{}^n)$$

$$= M(\mathbf{x}_t{}^n) + \mathbf{M}^n e_a{}^n + \ldots$$

where M is the non-linear model

and $\mathbf{M}^n$ is called the tangent linear model

$$\mathbf{x}_t{}^{n+1} = M(\mathbf{x}_t{}^n) - e_m$$

Subtract: $e_b{}^{n+1} = \mathbf{M}^n e_a{}^n + e_m$

The forecast error covariance is :

$$\mathbf{B}(t_{n+1}) = \langle (e_b{}^{n+1})(e_b{}^{n+1})^{\mathrm{T}} \rangle$$

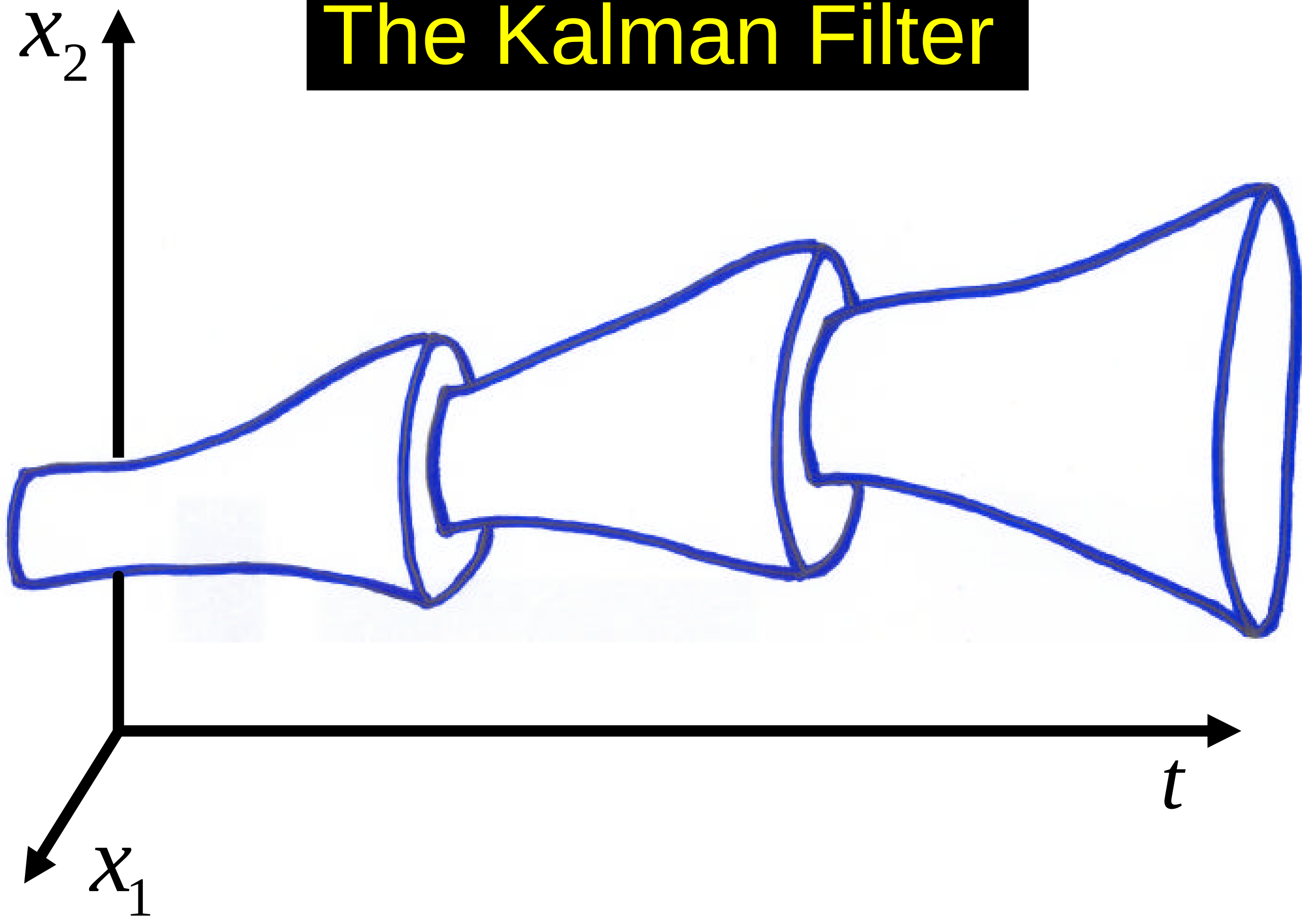
$$= \mathbf{M}(t_n)\mathbf{A}(t_n)\mathbf{M}^{\mathrm{T}}(t_n) + \mathbf{Q}(t_n)$$

where $\mathbf{Q} = \langle e_m e_m{}^{\mathrm{T}} \rangle$

$\mathbf{M}^{\mathrm{T}}$ is the adjoint of the tangent linear model.

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$$

The Kalman Filter



Remarks

- In OI (and 3d-VAR) isolated observation given more weight than observations close together (forecast errors have large correlations at nearby observation points).
- When several observations are close together calculation of weights may be ill-posed. Therefore combine into a “super observation”.

Extended Kalman Filter

- Assumes the model is *non-linear* and imperfect.
- The tangent linear model depends on the state and on time.
- Could be a “gold standard” for data assimilation, but very expensive to implement because of the very large dimension of the state space (~ 10^6 – 10^7 for NWP models).

Ensemble Kalman Filter

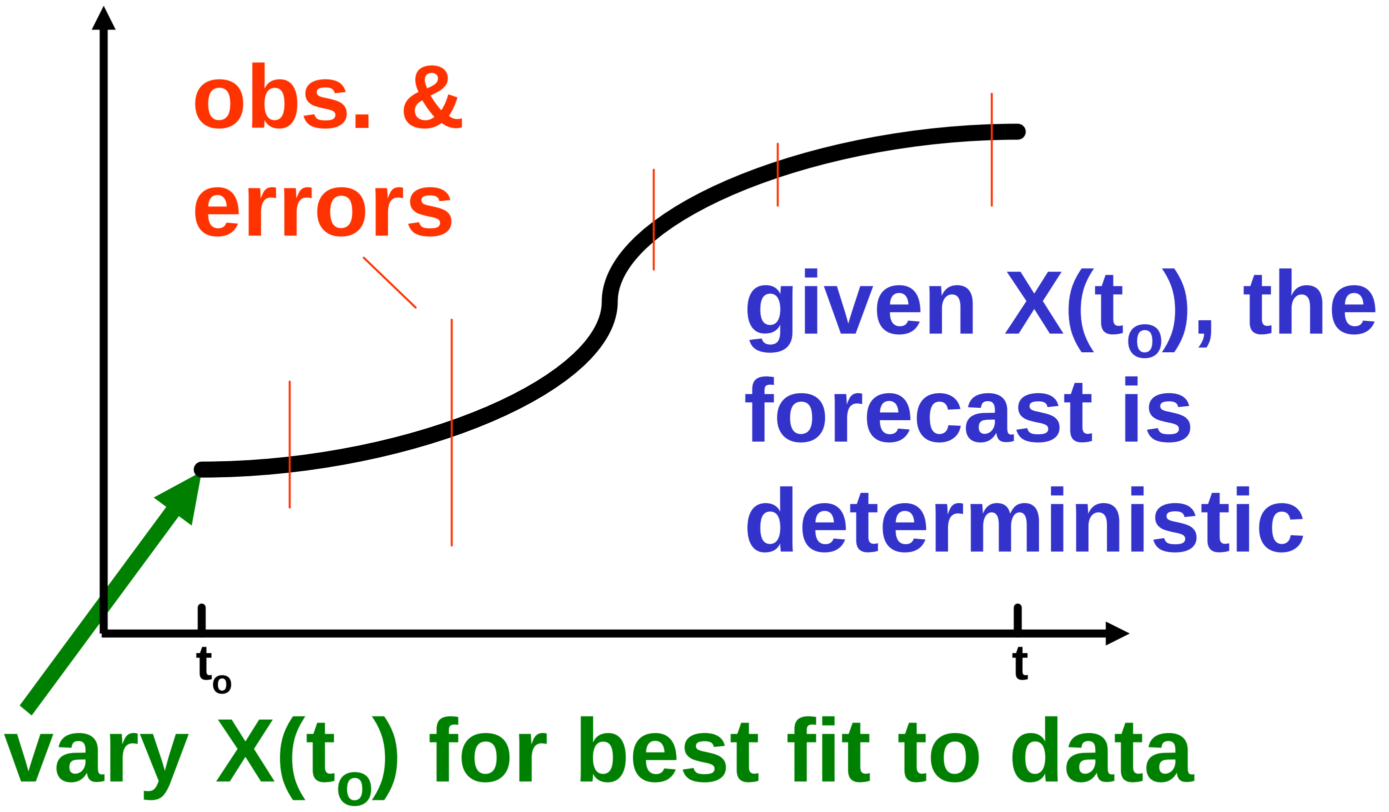
- Carry forecast error covariance matrix forward in time by using ensembles of forecasts:

$$\mathbf{B} \approx \frac{1}{K-1} \sum_{k \neq 1}^{K} (\mathbf{x}_k^f - <\mathbf{x}^f>)(\mathbf{x}_k^f - <\mathbf{x}^f>)^{\mathrm{T}}$$

- Only ~ 10 + forecasts needed.
- Does not require computation of tangent linear model and its adjoint.
- Does not require linearization of evolution of forecast errors.
- Fits in neatly into ensemble forecasting.

4d-Variational Assimilation

4D Variational Data Assimilation



4d-Variational Assimilation

$$J(\mathbf{x}(t_0)) = \frac{1}{2}\sum_{i=0}^{N}[\mathbf{y}_i - H(\mathbf{x}_i)]^{\mathrm{T}}\mathbf{R}_i^{-1}[\mathbf{y}_i - H(\mathbf{x}_i)]$$

$$+\frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}^b(t_0)]^{\mathrm{T}}\mathbf{B}_0^{-1}[\mathbf{x}(t_0) - \mathbf{x}^b(t_0)]$$

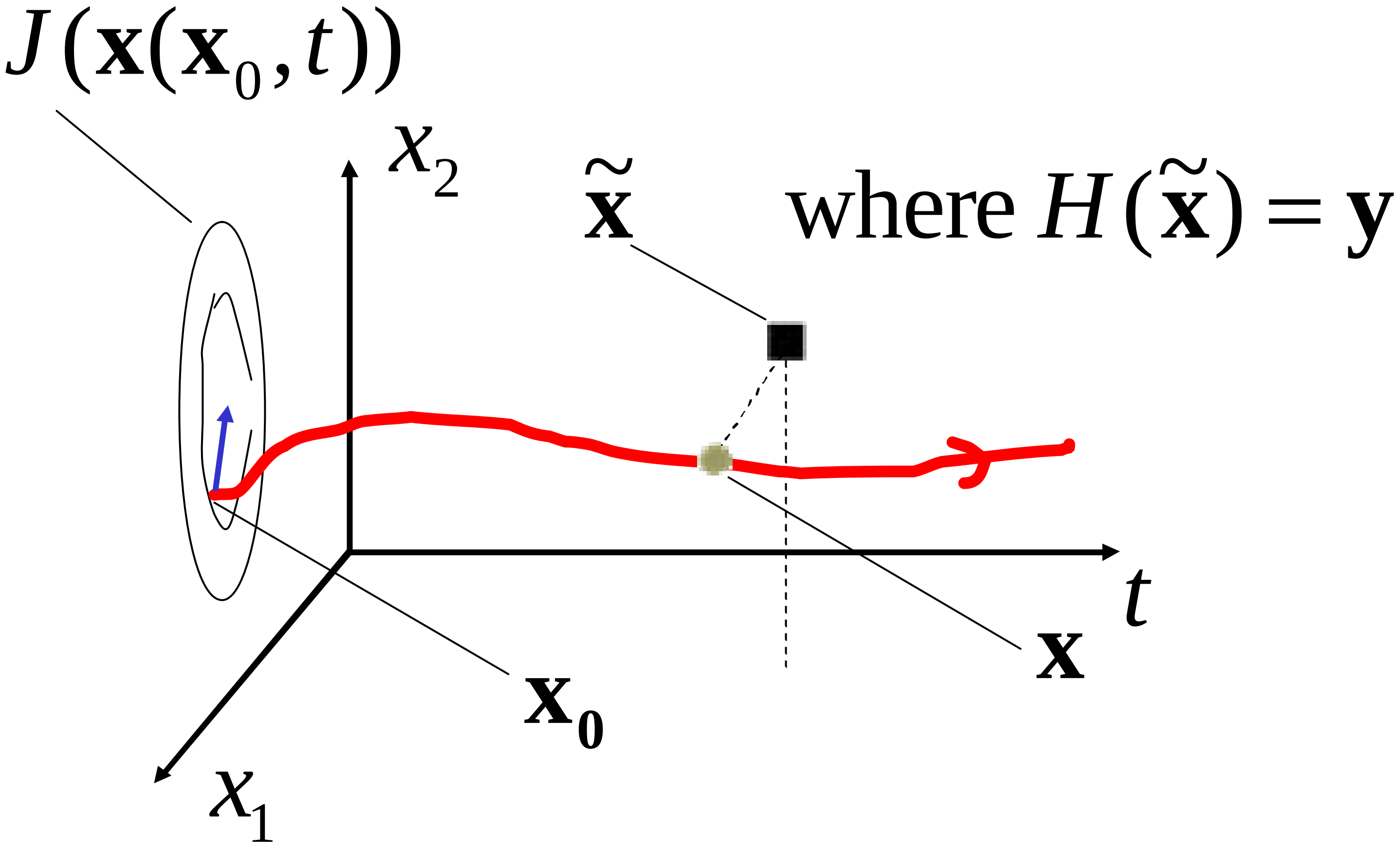
where $\mathbf{x}(t_i) = M_{0\to i}(\mathbf{x}(t_0))$ i.e. the model is treated as a strong constraint

Minimize the cost function by finding the gradient $\partial J / \mathbf{x}(t_0)$ ("Jacobian") with respect to the control variables in $\mathbf{x}(t_0)$

4d-VAR Continued

The 2nd term on the RHS of the cost function measures the distance to the background at the beginning of the interval. The term helps join up the sequence of optimal trajectories found by minimizing the cost function for the observations. The “analysis” is then the optimal trajectory in state space. Forecasts can be run from any point on the trajectory, e.g. from the middle.

4d-VAR For Single Observation at time t



Some Matrix Algebra

$J = J(\mathbf{x}(\mathbf{x}_0))$

Then $\dfrac{\partial J}{\partial \mathbf{x}_0} = \left(\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right)^{\mathrm{T}} \dfrac{\partial J}{\partial \mathbf{x}}$

adjoint of the model $M : \mathbf{x}_0 \mapsto \mathbf{x}$

Let J have the following form : $J = \mathbf{z}^{\mathrm{T}}(\mathbf{x})\mathbf{A}\mathbf{z}(\mathbf{x})$

Then it can be shown that $\dfrac{\partial J}{\partial \mathbf{x}} = \left(\dfrac{\partial \mathbf{z}}{\partial \mathbf{x}}\right)^{\mathrm{T}} \mathbf{A}\mathbf{z}$

Combining these results : $\dfrac{\partial J}{\partial \mathbf{x}_0} = \left(\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right)^{\mathrm{T}} \left(\dfrac{\partial \mathbf{z}}{\partial \mathbf{x}}\right)^{\mathrm{T}} \mathbf{A}\mathbf{z}$

4d-VAR for Single Observation

$$J(\mathbf{x}(\mathbf{x}_0)) = \frac{1}{2}[\mathbf{y} - H(\mathbf{x}(\mathbf{x}_0))]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x}(\mathbf{x}_0))]$$

obs. term only

By using results on slide "Some Matrix Algebra":

$$\frac{\partial J}{\partial \mathbf{x}_0} = -\mathbf{L}_{0\to t}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x}(\mathbf{x}_0))] \equiv -\mathbf{L}_{0\to t}^{\mathrm{T}}\mathbf{d}$$

where $\mathbf{L}_{0\to t}^{\mathrm{T}} = \left(\frac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right)^{\mathrm{T}} = \left(\frac{\partial M_{0\to t}(\mathbf{x}_0)}{\partial \mathbf{x}_0}\right)^{\mathrm{T}}$, adjoint of tangen

linear model

$$\mathbf{L}_{0\to t} = \mathbf{L}_{t_{n-1}\to t} \ldots \mathbf{L}_{t_1\to t_2}\mathbf{L}_{0\to t_1}$$

$$\therefore \mathbf{L}_{0\to t}^{\mathrm{T}} = \mathbf{L}_{0\to t_1}^{\mathrm{T}}\mathbf{L}_{t_1\to t_2}^{\mathrm{T}} \ldots \mathbf{L}_{t_{n-1}\to t}^{\mathrm{T}}$$

$\Rightarrow$ backward integration in

time of TLM

4d-VAR Procedure

- Choose $\mathbf{x}_0 \, , \mathbf{x}_0^b$ for example.
- Integrate full (non-linear) model forward in time and calculate $\mathbf{d}$ for each observation.
- Map $\mathbf{d}$ back to *t*=0 by backward integration of TLM, and sum for all observations to give the gradient of the cost function.
- Move down the gradient to obtain a better initial state (new trajectory “hits” observations more closely)
- Repeat until some STOP criterion is met.

note: not the most efficient algorithm

Comments

- 4d-VAR can also be formulated by the method of Lagrange multipliers to treat the model equations as a constraint. The adjoint equations that arise in this approach are the same equations we have derived by using the chain rule of partial differential equations.
- If model is perfect and B_0 is correct, 4d-VAR at final time gives same result as extended Kalman filter (but the covariance of the analysis is not available in 4d-VAR).
- 4d-VAR analysis therefore optimal over its time window, but less expensive than Kalman filter.

Incremental Form of 4d-VAR

- The 4d-VAR algorithm presented earlier is expensive to implement. It requires repeated forward integrations with the non-linear (forecast) model and backward integrations with the TLM.
- When the initial background (first-guess) state and resulting trajectory are accurate, an incremental method can be made much cheaper to run on a computer.

Incremental Form of 4d-VAR

The incremental form of the cost function is defined by

$$J(d\mathbf{x}_0) = \frac{1}{2}(d\mathbf{x}_0)^{\mathrm{T}}\mathbf{B}_0^{-1}(d\mathbf{x}_0)$$

where $d\mathbf{x}_0 = \mathbf{x}(t_0) - \mathbf{x}^b(t_0)$

$$+\frac{1}{2}\sum_{i=0}^{N}[\mathbf{y}_i - H(\mathbf{x}^f(t_i)) - \mathbf{H}_i\mathbf{L}(t_0,t_i)d\mathbf{x}_0]^{\mathrm{T}}\mathbf{R}_i^{-1}[\mathbf{y}_i - H(\mathbf{x}^f(t_i)) - \mathbf{H}_i\mathbf{L}(t_0,t_i)d\mathbf{x}_0]$$

Taylor series expansion about first-guess trajectory $\mathbf{x}^f(t_i)$

Minimization can be done in lower dimensional space

4D Variational Data Assimilation

- Advantages
  - consistent with the governing eqs.
  - implicit links between variables
- Disadvantages
  - very expensive
  - model is strong constraint

Some Useful References

- Atmospheric Data Analysis by R. Daley, Cambridge University Press.
- Atmospheric Modelling, Data Assimilation and Predictability by E. Kalnay, C.U.P.
- The Ocean Inverse Problem by C. Wunsch, C.U.P.
- Inverse Problem Theory by A. Tarantola, Elsevier.
- Inverse Problems in Atmospheric Constituent Transport by I.G. Enting, C.U.P.
- ECMWF Lecture Notes at www.ecmwf.int

END