Data Assimilation

Alan O'Neill
National Centre for Earth Observation
&
University of Reading

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Contents

- Motivation
- Univariate (scalar) data assimilation
- Multivariate (vector) data assimilation
  - Optimal Interpoletion (*BLUE*)
  - 3d-Variational Method
  - Kalman Filter
  - 4d-Variational Method
- Applications of data assimilation in earth system science

Motivation

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.

DARC

Why We Need Data Assimilation

- range of observations
- range of techniques
- different errors
- data gaps
- quantities not measured
- quantities linked

Preliminary Concepts

What We Want To Know

$\mathbf{x}(t)$ atmos. state vector

$\mathbf{s}(t)$ surface fluxes

$\mathbf{c}$ model parameters

$$\mathbf{X}(t) = (\mathbf{x}(t), \mathbf{s}(t), \mathbf{c})$$

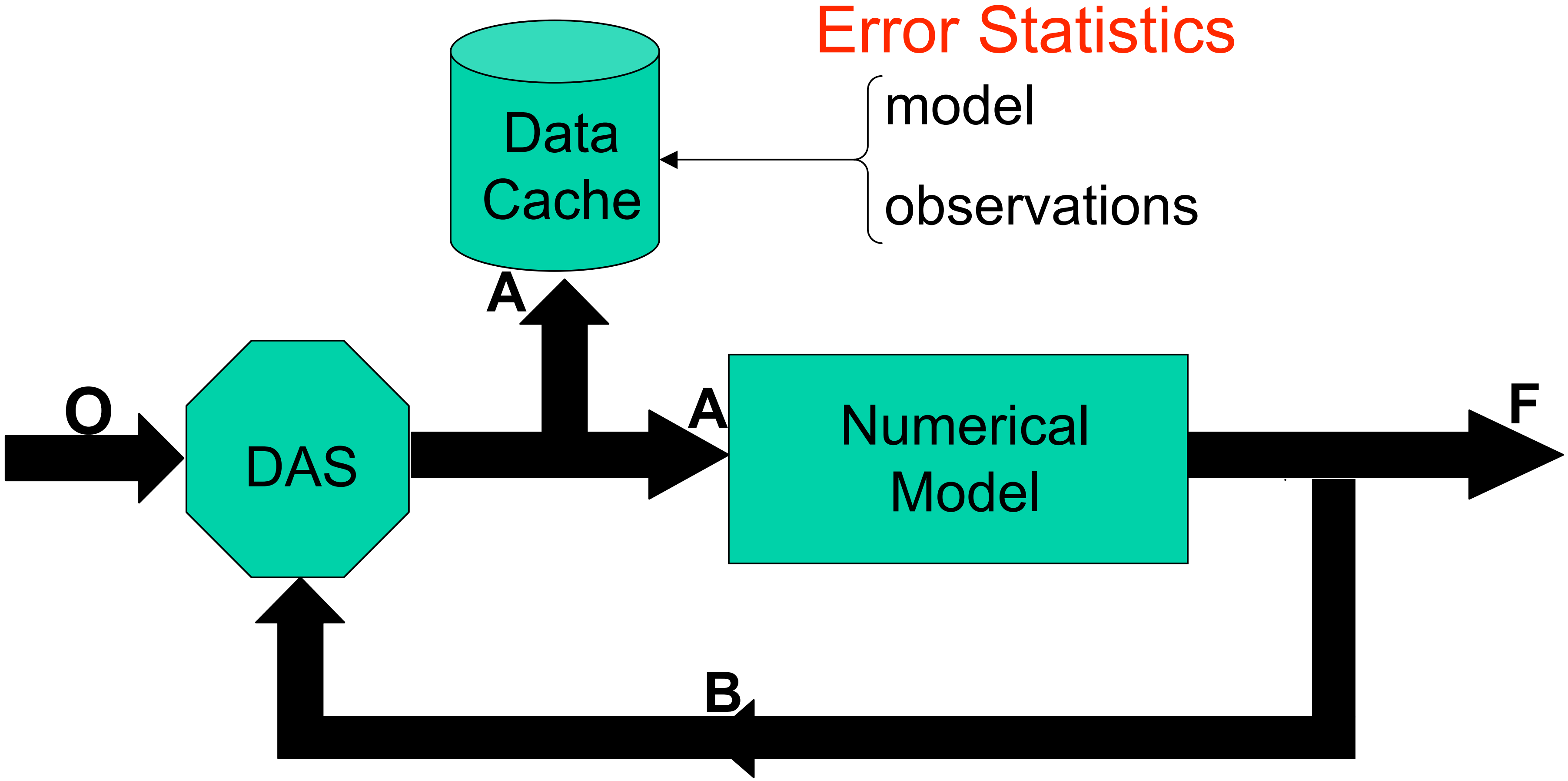
What We Also Want To Know

Errors in models

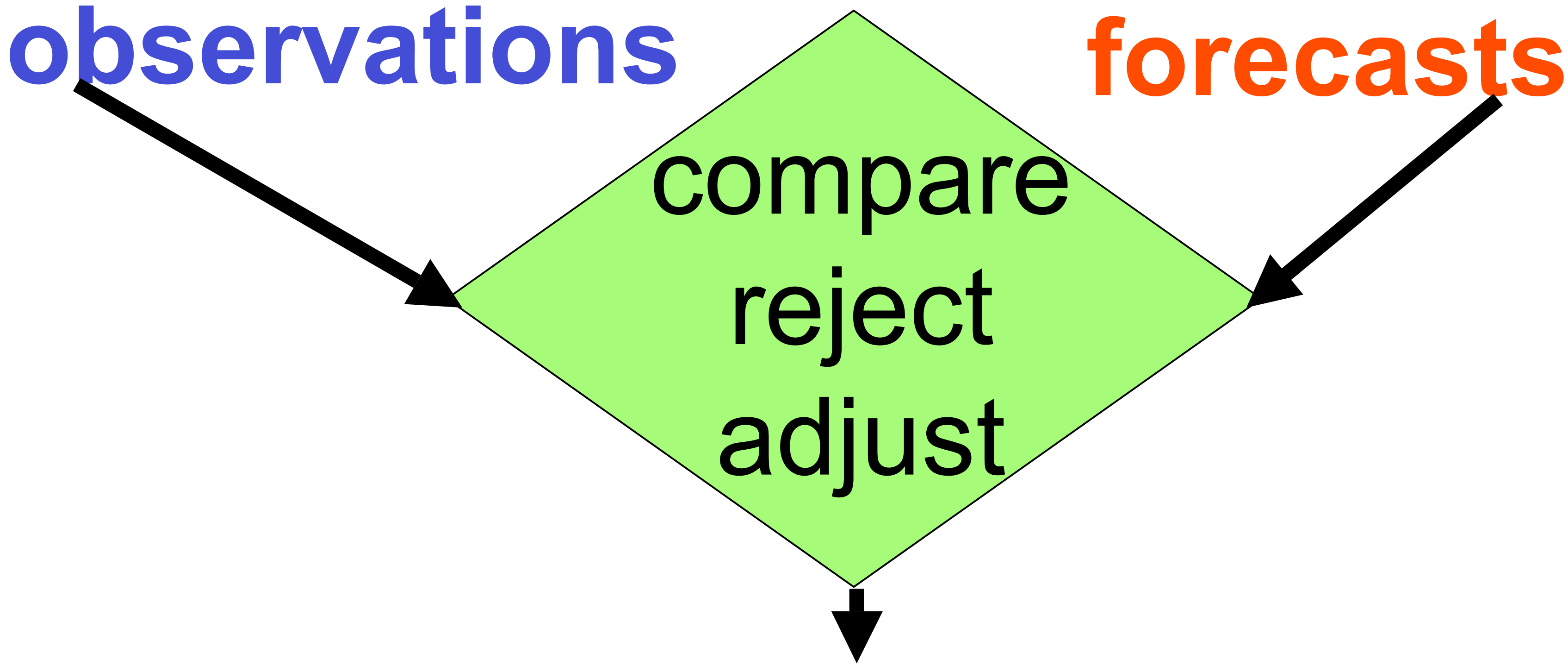
Errors in observations

What observations to make

DATA ASSIMILATION SYSTEM

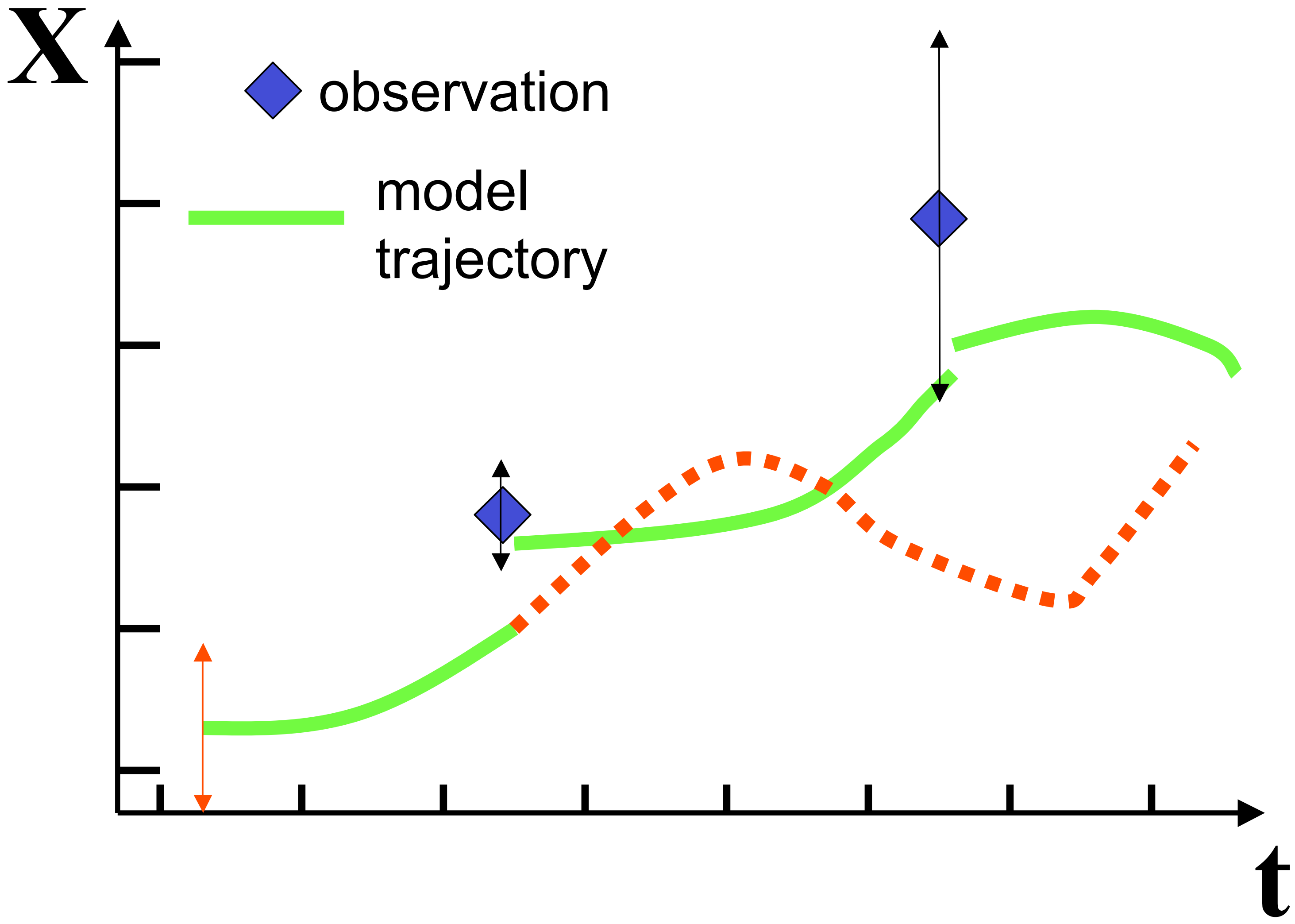


The Data Assimilation Process



estimates of state & parameters

➡ **errors in obs. & forecasts**


X
observation
model
trajectory
t

Basic Concept of Data Assimilation

- Information is accumulated in time into the model state and propagated to all variables.

What are the benefits of data assimilation?

- Quality control
- Combination of data
- Errors in data and in model
- Filling in data poor regions
- Designing observational systems
- Maintaining consistency
- Estimating unobserved quantities

DARC

Some Uses of Data Assimilation (state estimation, inverse modelling)

- Satellite retrievals
- Operational weather and ocean forecasting
- Seasonal weather forecasting
- Land-surface process
- Surface-flux estimation
- Model parameter estimation
- Global climate datasets
- Planning satellite measurements
- Evaluation of models and observations

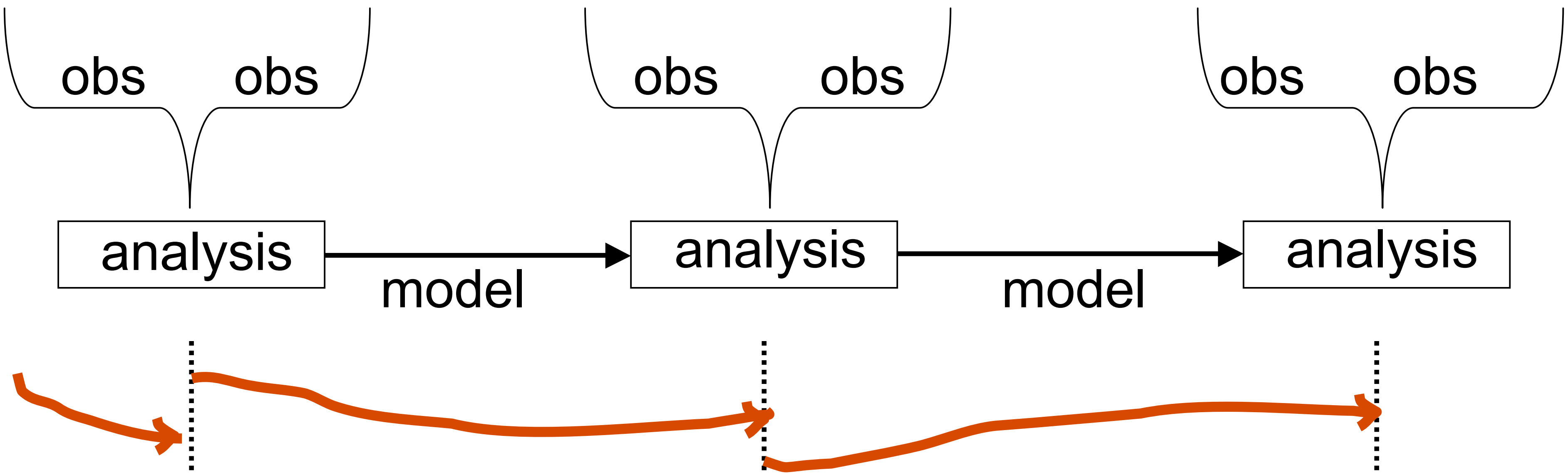
DARC

Types of Data Assimilation

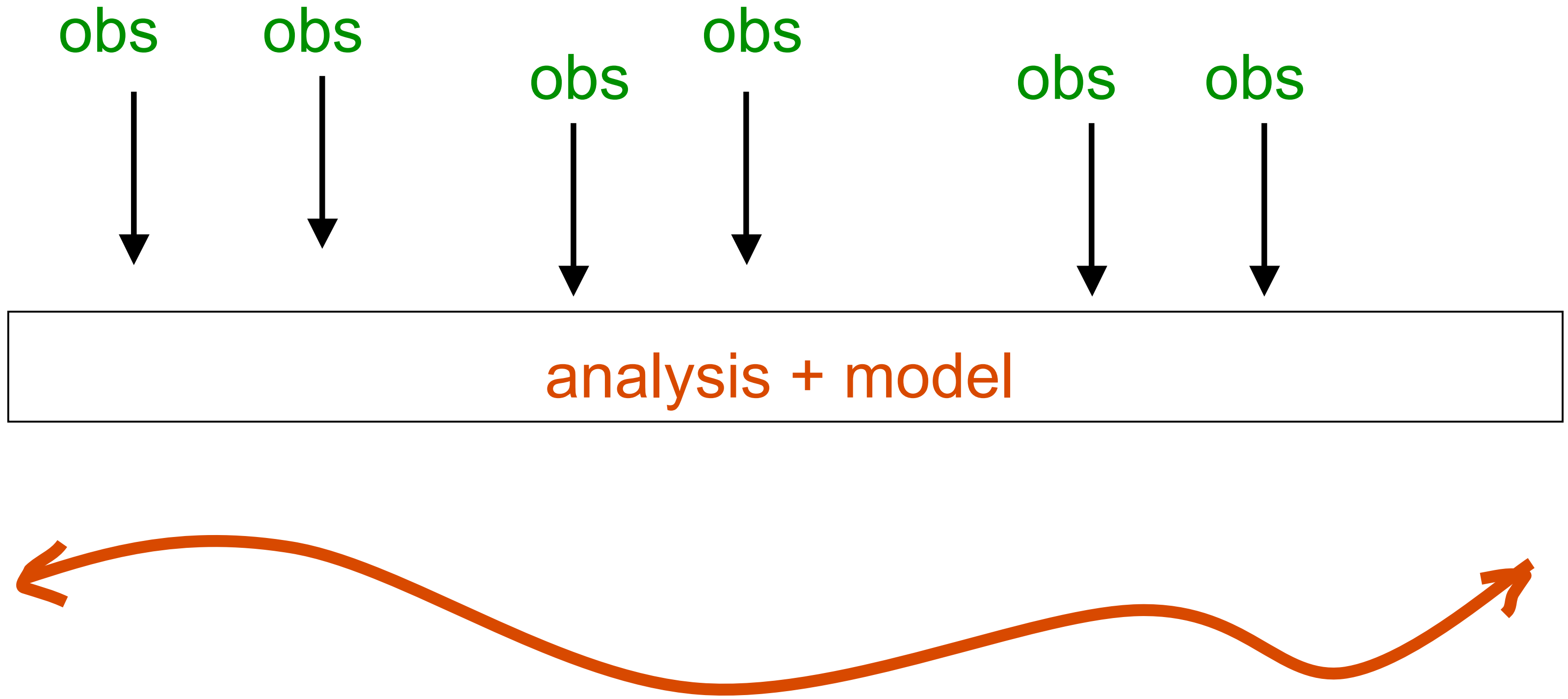
- Sequential
- Non-sequential (4D-variational)

- Intermittent
- Continuous

Sequential Intermittent Assimilation



Non-sequential Continuous Assimilation



Methods of Data Assimilation

- Optimal interpolation (or approx. to it)
- 3D variational method (3DVar)
- Kalman filter (with approximations)
- 4D-variational (4DVar)

DARC

Statistical Approach to Data Assimilation

DARC

Data Assimilation
Made Simple
(scalar case)

Least Squares Method (Minimum Variance)

$$T_1 = T_t + \varepsilon_1$$

$$T_2 = T_t + \varepsilon_2$$

$$<\varepsilon_1> = <\varepsilon_2> = 0$$

$$<(\varepsilon_1)^2> = \sigma_1^{\ 2}$$

$$<(\varepsilon_2)^2> = \sigma_2^{\ 2}$$

$<\varepsilon_1\varepsilon_2> = 0$, the two measurements are uncorrelated

Estimate T_{t} as a linear combination of the observations

$$T_a = a_1 T_1 + a_2 T_2$$

The analysis should be unbiased: $<T_a> = <T_t>$

$$\Rightarrow a_1 + a_2 = 1$$

Least Squares Method Continued

Estimate T_a by minimizing its mean squared error :

$$\sigma_a{}^2 =< (T_a - T_t)^2 >=< (a_1(T_1 - T_t) + a_2(T_2 - T_t))^2 >$$

$$=< (a_1\varepsilon_1 + a_2\varepsilon_1)^2 >= a_1\sigma_1{}^2 + a_2\sigma_2{}^2$$

subject to the constraint $\quad a_1 + a_2 = 0$

Least Squares Method Continued

$$a_1 = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} \qquad a_2 = \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

$$\Rightarrow \frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

The precision of the analysis is the sum of the precisions of the measurements. The analysis therefore has higher precision than any single measurement (if the statistics are correct).

Maximum Likelihood Estimate

- Obtain or assume probability distributions for the errors
- The best estimate of the state is chosen to have the greatest probability, or maximum likelihood
- If errors normally distributed,unbiased and uncorrelated, then states estimated by minimum variance and maximum likelihood are the same

Maximum Likelihood Approach (Bayesian Derivation)

Assume we have already made an observation T_1, (the background forecast in data assimilation).
Then the probability distribution of the truth T for Gaussian error statistics is :

$$p(T) \propto \exp\left(-\frac{(T - T_1)^2}{2\sigma_1^{\ 2}} \right) \quad \text{the prior pdf.}$$

Maximum Likelihood Continued

Bayes's formula for the posterior pdf given the observation T_2 is:

$$p(T \mid T_2) = \frac{p(T_2 \mid T)\, p(T)}{p(T_2)} \propto \exp\left(- \frac{(T_2 - T)^2}{2\sigma_2^{\,2}} \right) \exp\left(- \frac{(T - T_1)^2}{2\sigma_1^{\,2}} \right)$$

since $p(T_2)$ is normalising factor independent of T.

Maximize the posterior pdf (or ln) to estimate the truth.

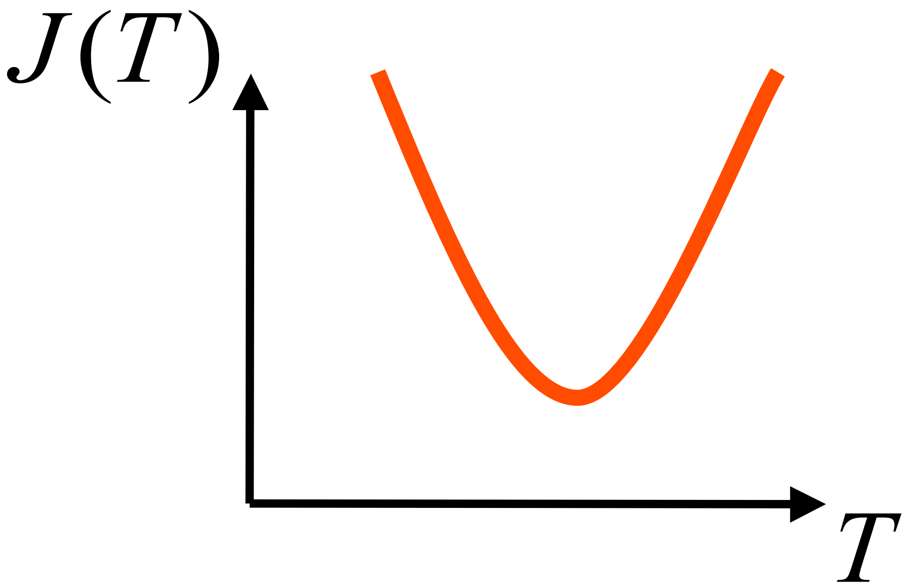
We get the same answer as minimizing the cost function.

Equivalence holds for multi-dimensional case (for Gaussian statistics).

Variational Approach

$$J(T) = \frac{1}{2}\left[\frac{(T-T_1)^2}{\sigma_1^{\ 2}} + \frac{(T-T_2)^2}{\sigma_2^{\ 2}}\right]$$

T_a is the value of T for which $\partial J / \partial T = 0$



Simple Sequential Assimilation

Let $T_1 = T_b$ $T = T_o$

$T_a = T_b + W(T_o - T_b)$ where $(T_o - T_b)$ is the "innovation".

The optimal weight W is given by:

$W = \sigma_b{}^2(\sigma_b{}^2 + \sigma_o{}^2)^{-1}$, and the analysis error variance is:

$\sigma_a{}^2 = (1 - W)\sigma_b{}^2$

Comments

- The analysis is obtained by adding first guess to the innovation.
- Optimal weight is background error variance multiplied by inverse of total variance.
- Precision of analysis is sum of precisions of background and observation.
- Error variance of analysis is error variance of background *reduced* by (1- optimal weight).

Simple Assimilation Cycle

- Observation used once and then discarded.
- Forecast phase to update T_b and σ_b^2
- Analysis phase to update T_a and σ_a^2
- Obtain background as

$$T_b(t_{i+1}) = M[T_a(t_i)]$$

- Obtain variance of background as

$$\sigma_b^2(t_{i+1}) = \sigma_b^2(t_i) \qquad \text{alternatively} \qquad \sigma_b^2(t_{i+1}) = a\sigma_a^2(t_i)$$

Simple Kalman Filter

Analysis step as before.

$$T_t(t_{i+1}) = M[T_t(t_i)] - \varepsilon_m, \quad Q^2 = < \varepsilon_m{}^2 > \qquad \text{(model not biased!)}$$

Then $\varepsilon_{b,i+1} = (T_b - T_t)_{i+1} = M(T_{a,i}) - M(T_{t,i}) + \varepsilon_m$

$$= \mathrm{M}\varepsilon_{a,i} + \varepsilon_m \quad \text{where } \mathrm{M} = \partial M / \partial T$$

Forecast background error covariance is:

$$(\sigma_{b,i+1})^2 = < (\varepsilon_{b,i+1})^2 > = \mathrm{M}^2 (\sigma_{a,i})^2 + Q^2$$

Multivariate Data Assimilation

Multivariate Case

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

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

State Vectors

$\mathbf{x}$ state vector (column matrix)

$\mathbf{x}_t$ true state

$\mathbf{x}_b$ background state

$\mathbf{x}_a$ analysis, estimate of $\mathbf{x}_t$

Ingredients of Good Estimate of the State Vector (“analysis”

- Start from a good “first guess” (forecast from previous good analysis)
- Allow for errors in observations and first guess (give most weight to data you trust)
- Analysis should be smooth
- Analysis should respect known physical laws

Some Useful Matrix Properties

Transpose of a product: $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$

Inverse of a product: $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Inverse of a transpose: $(\mathbf{A}^{\mathrm{T}})^{-1} = (\mathbf{A}^{-1})^{\mathrm{T}}$

Positive definiteness for symmetrix matrix $\mathbf{A}$:

$\forall\ \mathbf{x}$, the scalar $\mathbf{xAx}^{\mathrm{T}} > 0$, unless $\mathbf{x} = \mathbf{0}$.

(this property is conserved through inversion)

Observations

- Observations are gathered into an observation vector $\mathbf{y}$, called the observation vector.
- Usually fewer observations than variables in the model; they are irregularly spaced; and may be of a different kind to those in the model.
- Introduce an observation operator to map from model state space to observation space.

$$\mathbf{x} \rightarrow H(\mathbf{x})$$

Errors

Variance becomes Covariance Matrix

- Errors in x_i are often correlated
  - spatial structure in flow
  - dynamical or chemical relationships
- Variance for scalar case becomes Covariance Matrix for vector case COV
- Diagonal elements are the variances of x_i
- Off-diagonal elements are covariances between x_i and x_j
- Observation of x_i affects estimate of x_j

The Error Covariance Matrix

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

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

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

Background Errors

- They are the estimation errors of the background state:

$$\varepsilon_{\mathrm{b}} = \mathbf{x}_{\mathrm{b}} - \mathbf{x}_{\mathrm{t}}$$

- average (bias) $< \varepsilon_{\mathrm{b}} >$
- covariance

$$\mathbf{B} = <(\varepsilon - <\varepsilon_{\mathrm{b}}>)(\varepsilon - <\varepsilon_{\mathrm{b}}>)^{\mathrm{T}}>$$

Observation Errors

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

$$\varepsilon_\mathrm{o} = \mathbf{y} - H(\mathbf{x}_\mathrm{t}) \qquad\qquad < \varepsilon_\mathrm{o} >$$

$$\mathbf{R} =< (\varepsilon_\mathrm{o} - < \varepsilon_\mathrm{o} >)(\varepsilon_\mathrm{o} - < \varepsilon_\mathrm{o} >)^\mathrm{T} >$$

Control Variables

- We may not be able to solve the analysis problem for all components of the model state (e.g. cloud-related variables, or need to reduce resolution)
- The work space is then not the model space but the sub-space in which we correct $\mathbf{x}_\mathrm{b}$, called control-variable space

$$\mathbf{x}_\mathrm{a} = \mathbf{x}_\mathrm{b} + \delta\mathbf{x}$$

Innovations and Residuals

- Key to data assimilation is the use of differences between observations and the state vector of the system
- We call $\mathbf{y} - H(\mathbf{x}_\mathrm{b})$ the innovation
- We call $\mathbf{y} - H(\mathbf{x}_\mathrm{a})$ the analysis residual

Give important information

Analysis Errors

- They are the estimation errors of the analysis state that we want to minimize.

$$\varepsilon_{\mathrm{a}} = \mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{t}}$$

Covariance matrix $\mathbf{A}$

Using the Error Covariance Matrix

Recall that an error covariance matrix $\mathbf{C}$ for the error in $\mathbf{x}$ has the form:

$$\mathbf{C} = \langle \varepsilon\varepsilon^{\mathrm{T}} \rangle$$

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

$$\mathbf{C}_y = \mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}}$$

BLUE Estimator

- The BLUE estimator 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 analysis error covariance matrix is:

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

- Note that:

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

Statistical Interpolation with Least Squares Estimation

- Called *Best Linear Unbiased Estimator (BLUE).*
- Simplified versions of this algorithm yield the most common algorithms used today in meteorology and oceanography.

Assumptions Used in *BLUE*

- Linearized observation operator:

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

- $\mathbf{B}$ and $\mathbf{R}$ are positive definite.
- Errors are unbiased:

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

- Errors are uncorrelated:

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

- Linear analysis: corrections to background depend linearly on (background – obs.).
- Optimal analysis: minimum variance estimate.

Optimal Interpolation

observation

observation operator

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H(\mathbf{x}_b))$$

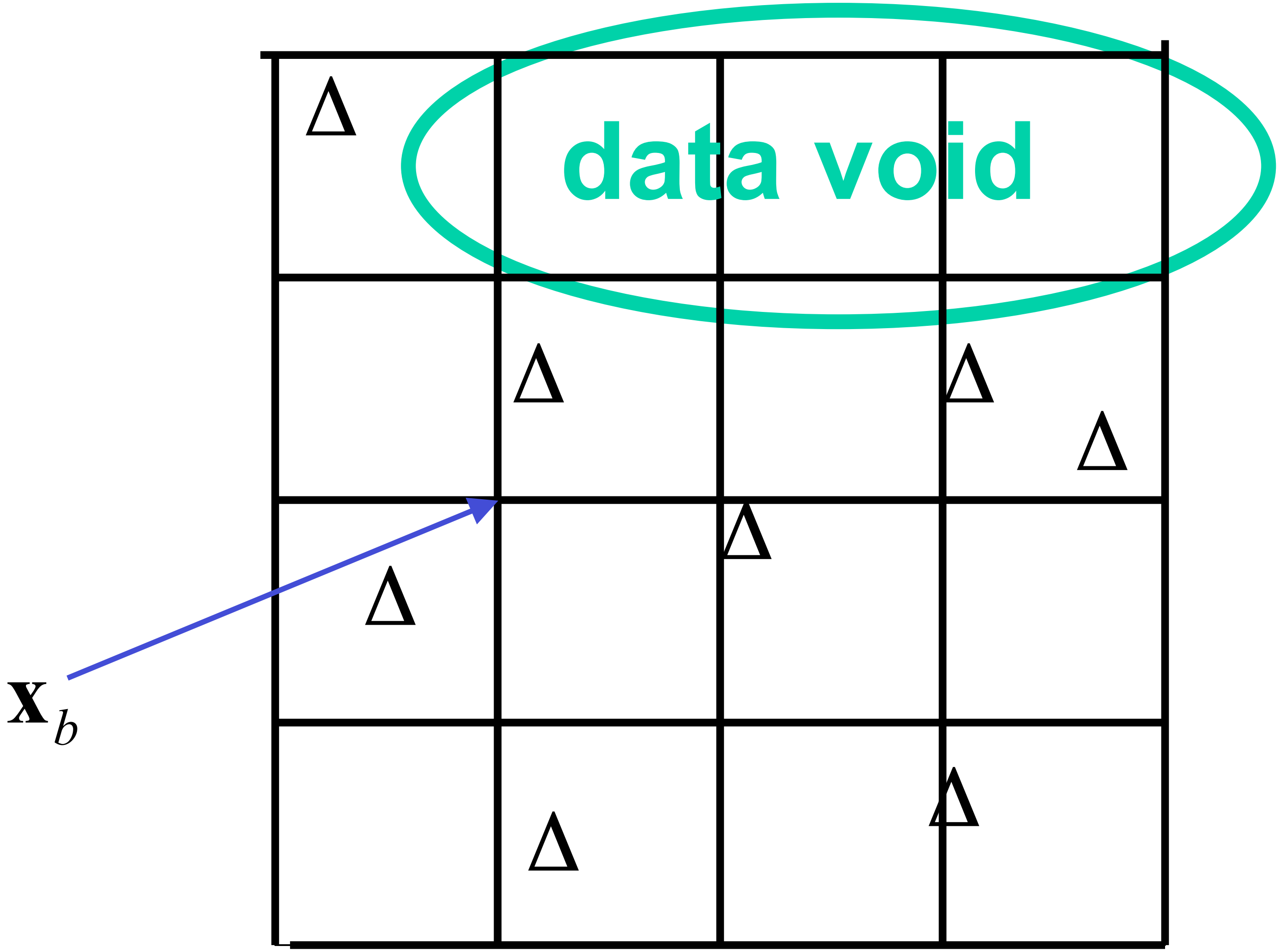
"analysis"

"background" (forecast)

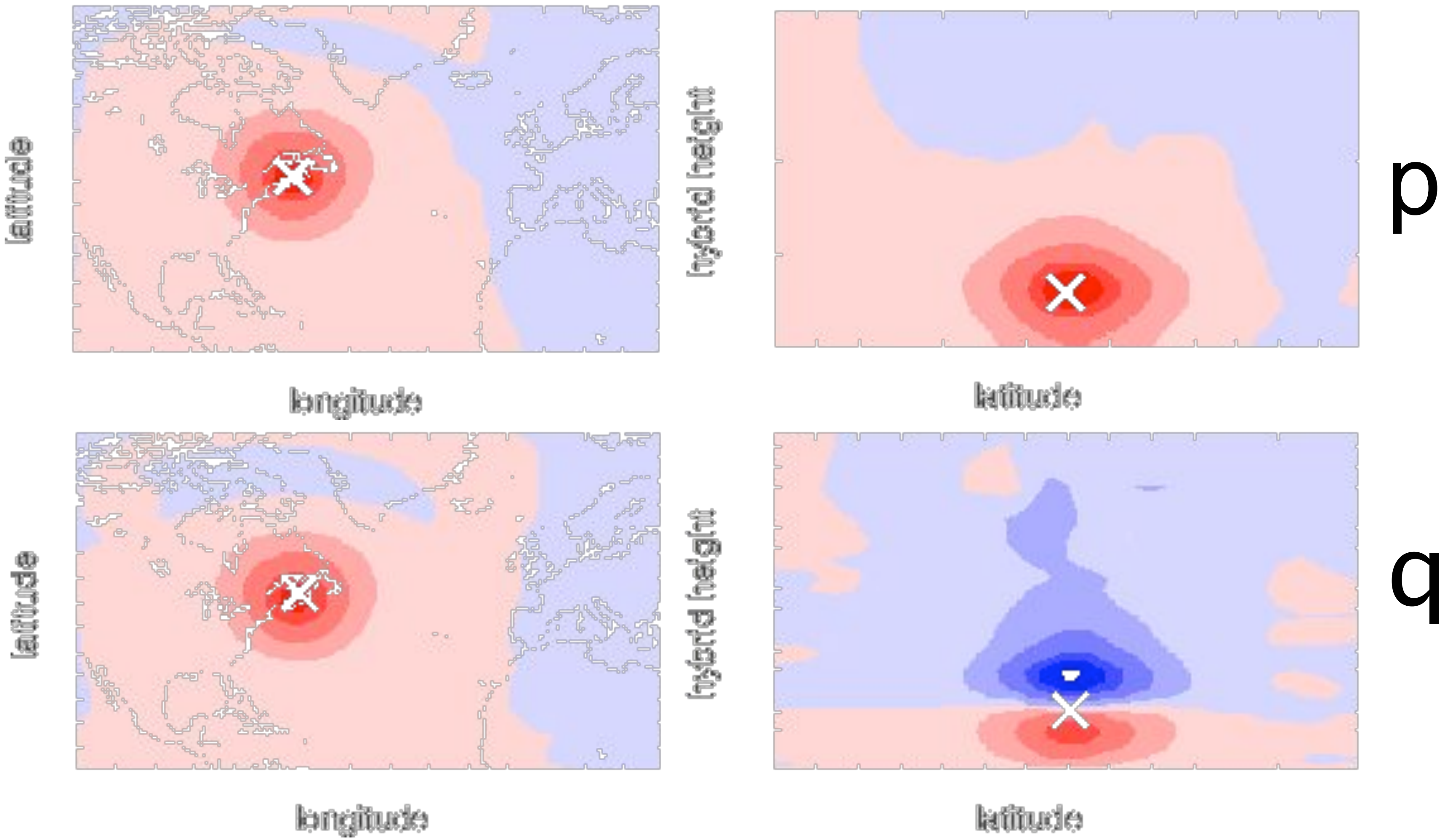
- **linearity** $H \rightarrow \mathrm{H}$
- **matrix inverse**
- **limited area**

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

$$\Delta = \mathbf{y} - H(\mathbf{x}_b) \longleftarrow$$ **at obs. point**

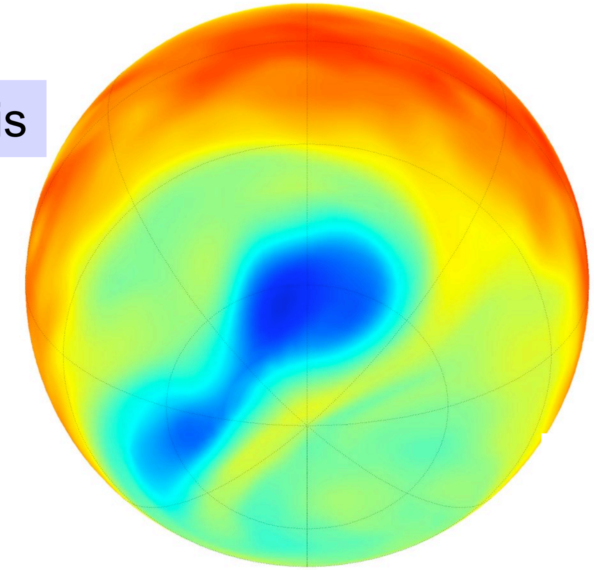


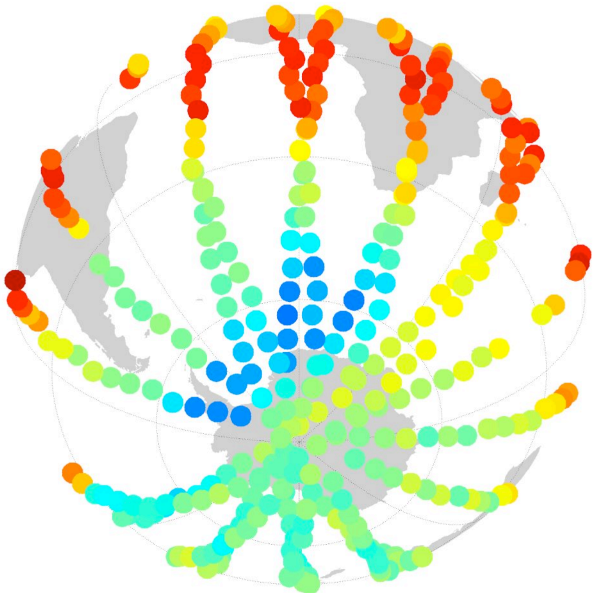
Spreading of Information from Single Pressure Obs.



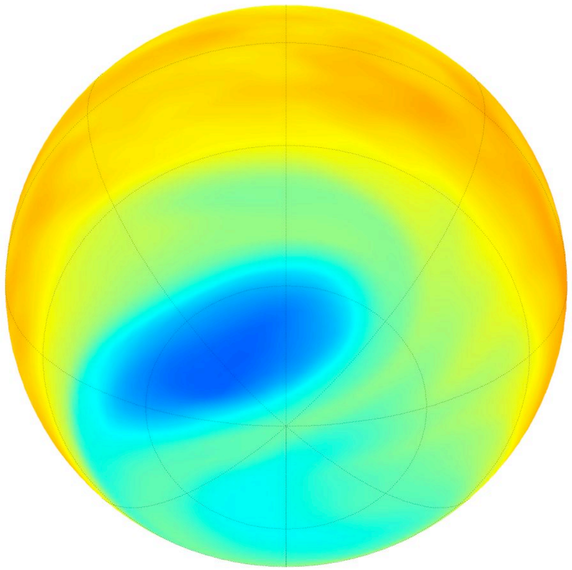
Ozone at 10hPa, 12Z 23rd Sept 2002

Analysis





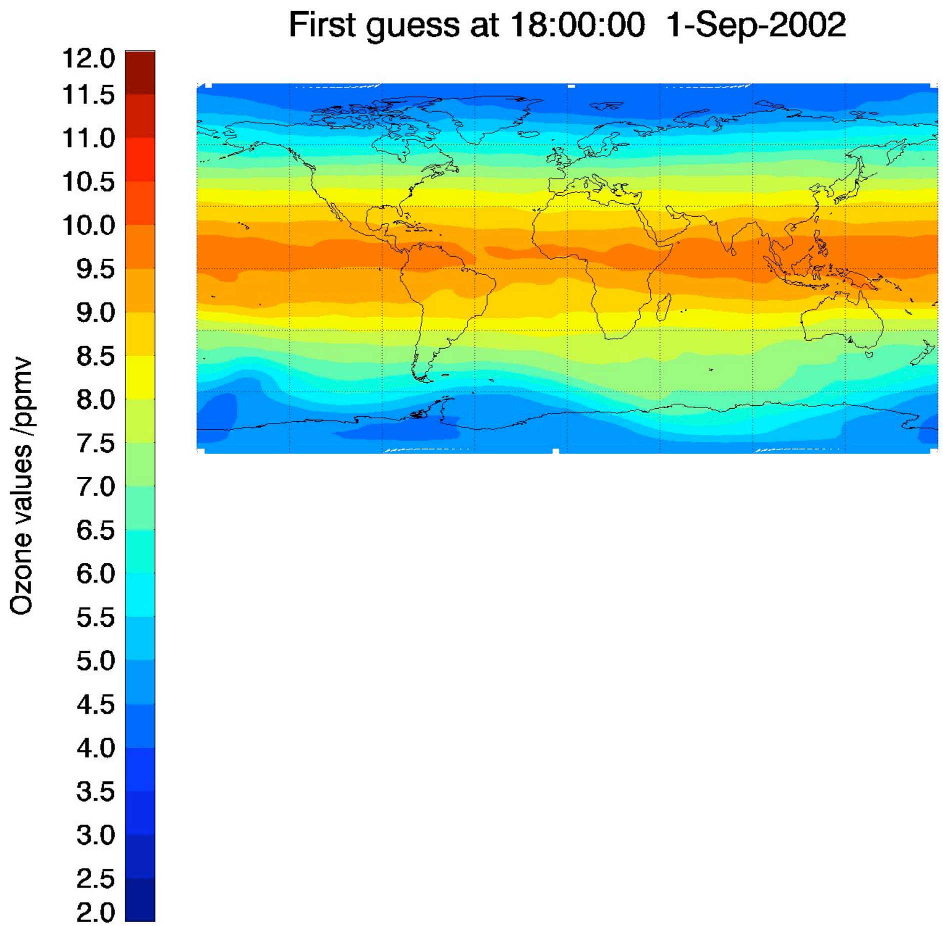
MIPAS observations



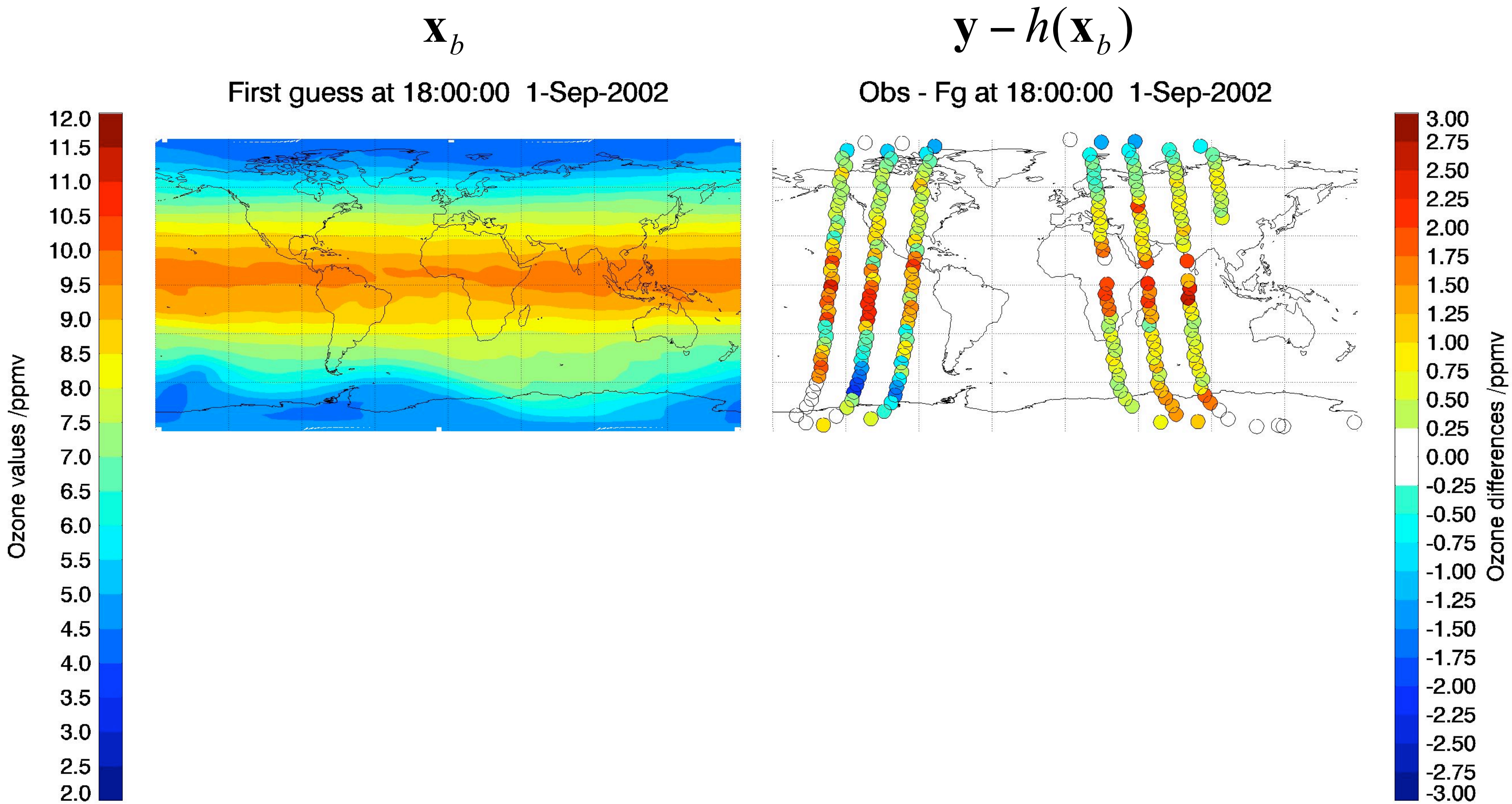
6 day model forecast

3D variational data assimilation - ozone at 10hPa

$\mathbf{x}_b$

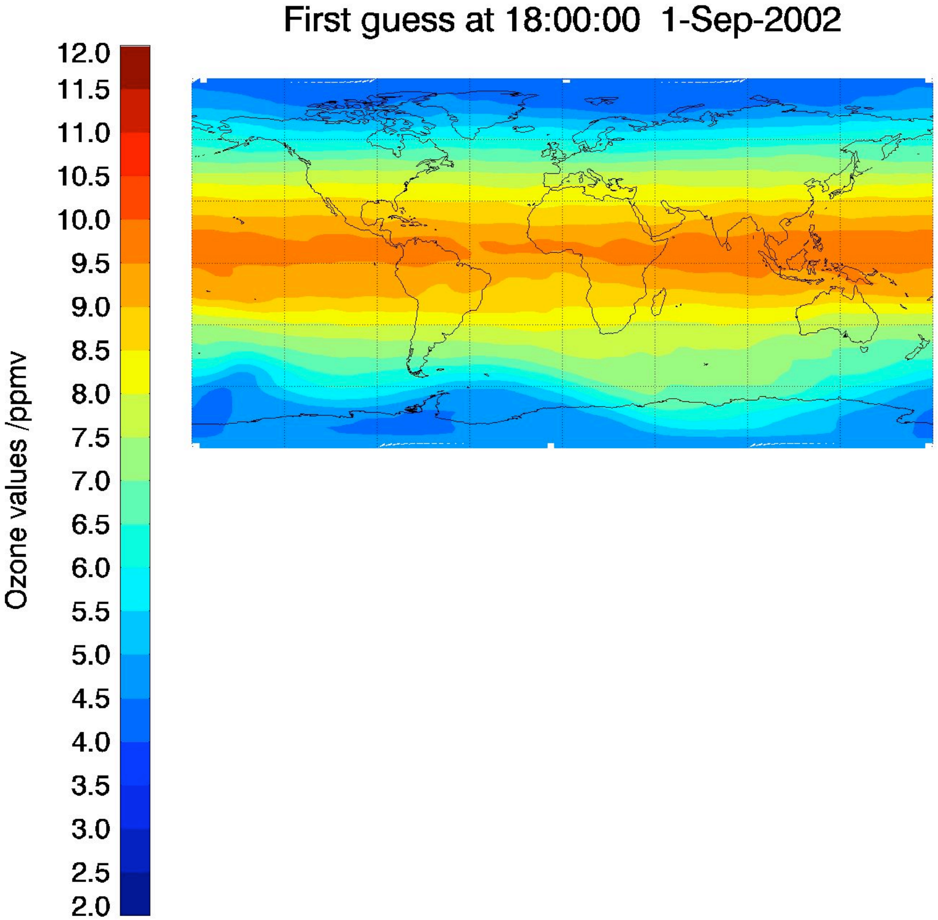


3D variational data assimilation - ozone at 10hPa

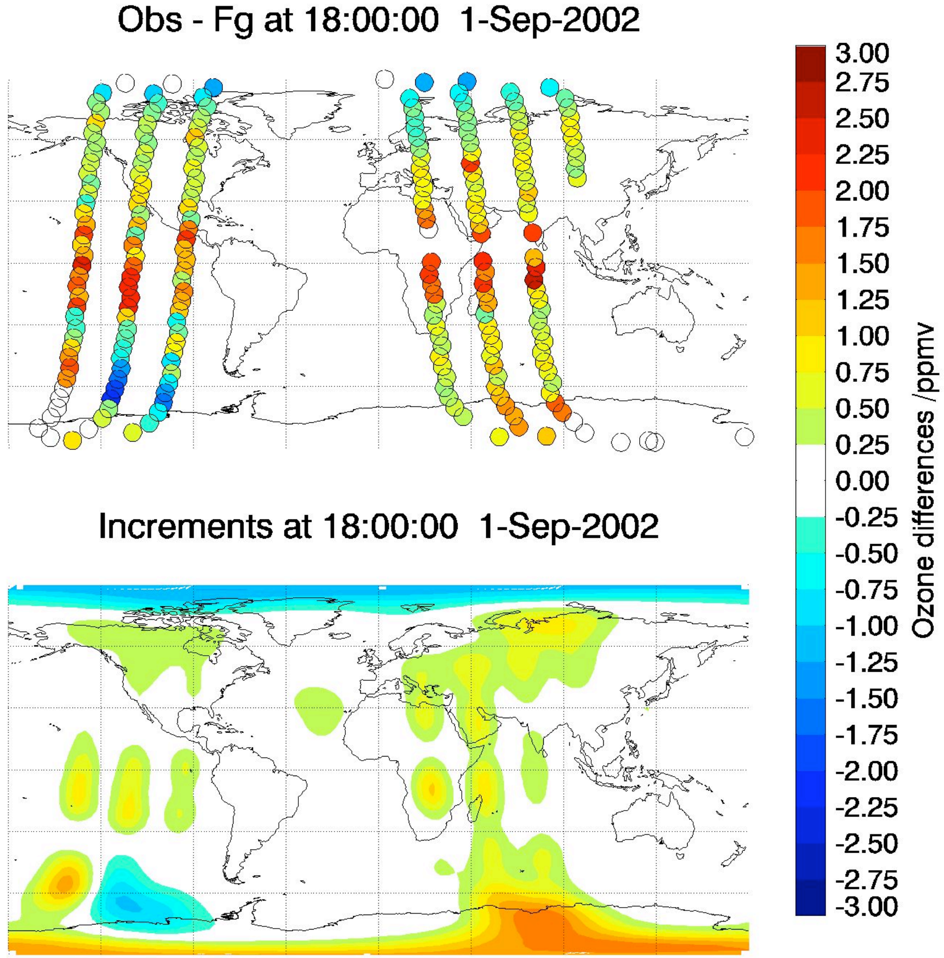


3D variational data assimilation - ozone at 10hPa

$\mathbf{x}_b$



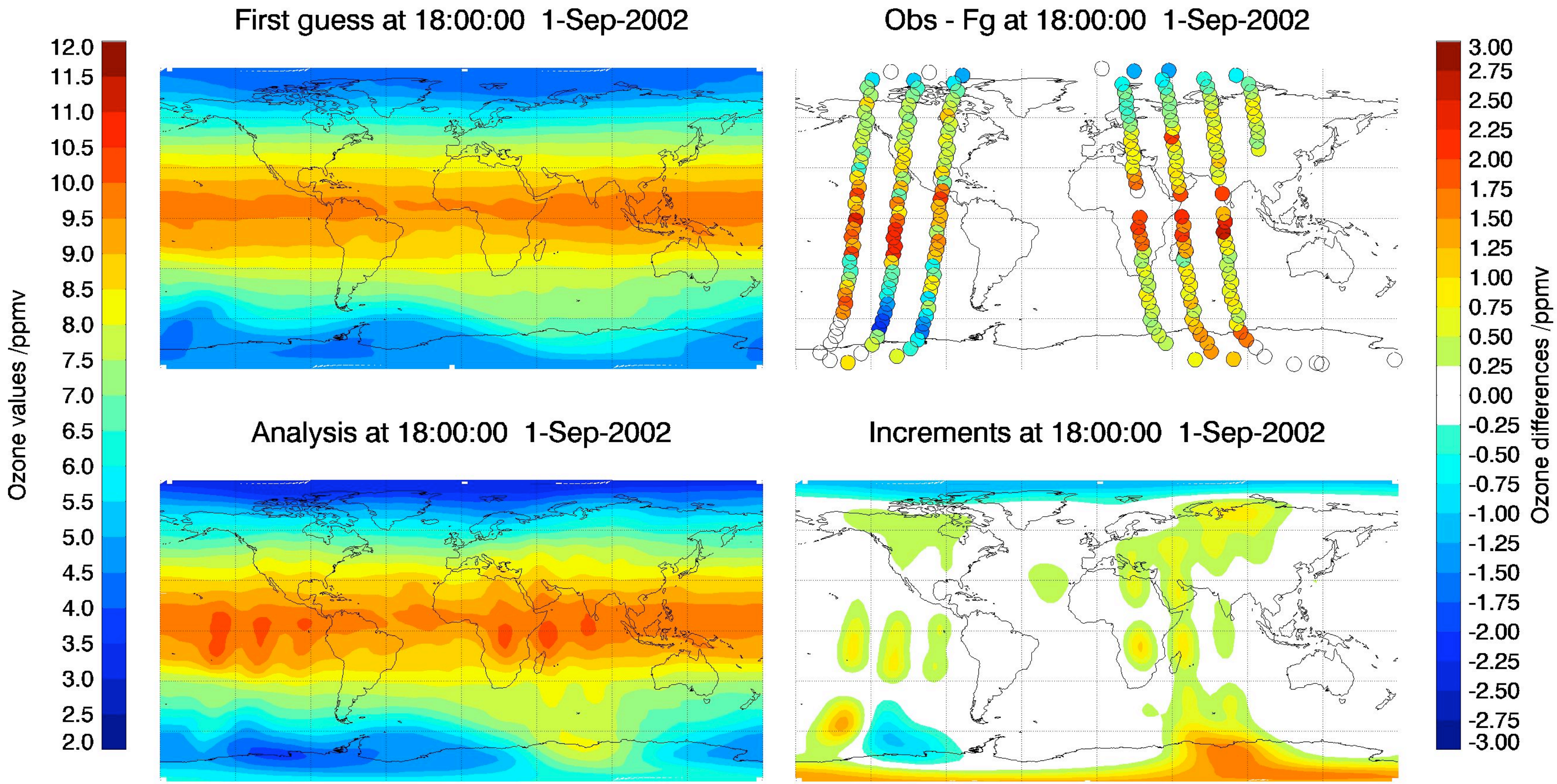
$\mathbf{y} - h(\mathbf{x}_b)$



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

The data assimilation cycle: ozone at 10hPa

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



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

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:

$$B_{ij} \propto \sigma_i \sigma_j \exp(-d_{ij}/L)$$

horiz. fn x vert. fn

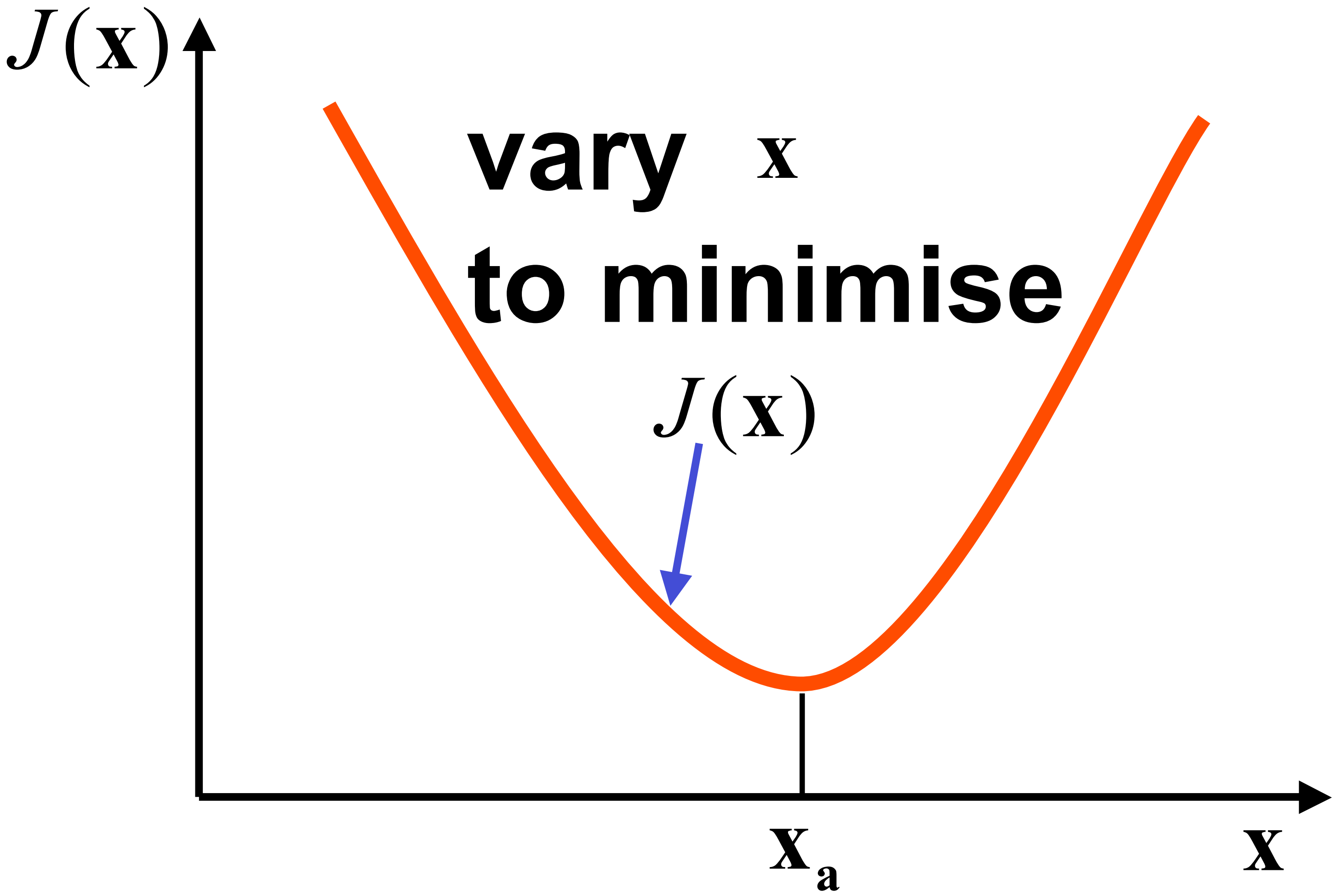
- Error growth in short-range forecasts “verifying” at the same time (NMC method)

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

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

3d-Variational Data Assimilation

Variational Data Assimilation



Equivalent Variational Optimization Problem

- BLUE analysis can be obtained by minimizing a cost (penalty, performance) function:

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

$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x})$$

$$\mathbf{x}_a = \min J$$

- The analysis $\mathbf{x}_a$ is optimal (closest in least-squares sense to $\mathbf{x}_t$).
- If the background and observation errors are Gaussian, then $\mathbf{x}_a$ is also the *maximum likelihood estimator*.

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.

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

Maximum Probability or Likelihood

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

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

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

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

where b, o, and a are normalizing factors.

- Maximum probability estimate minimizes

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

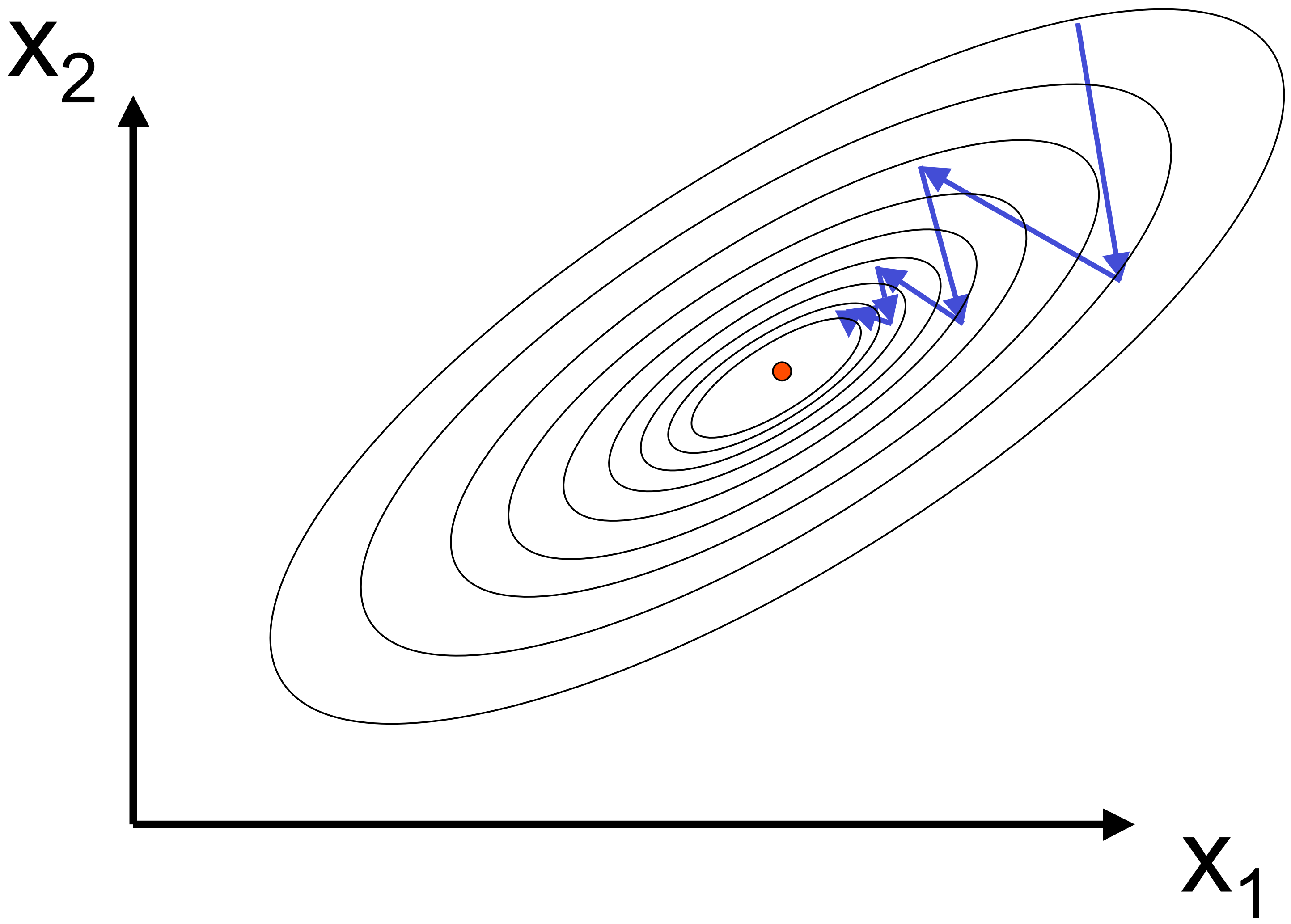
Comments

- 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.

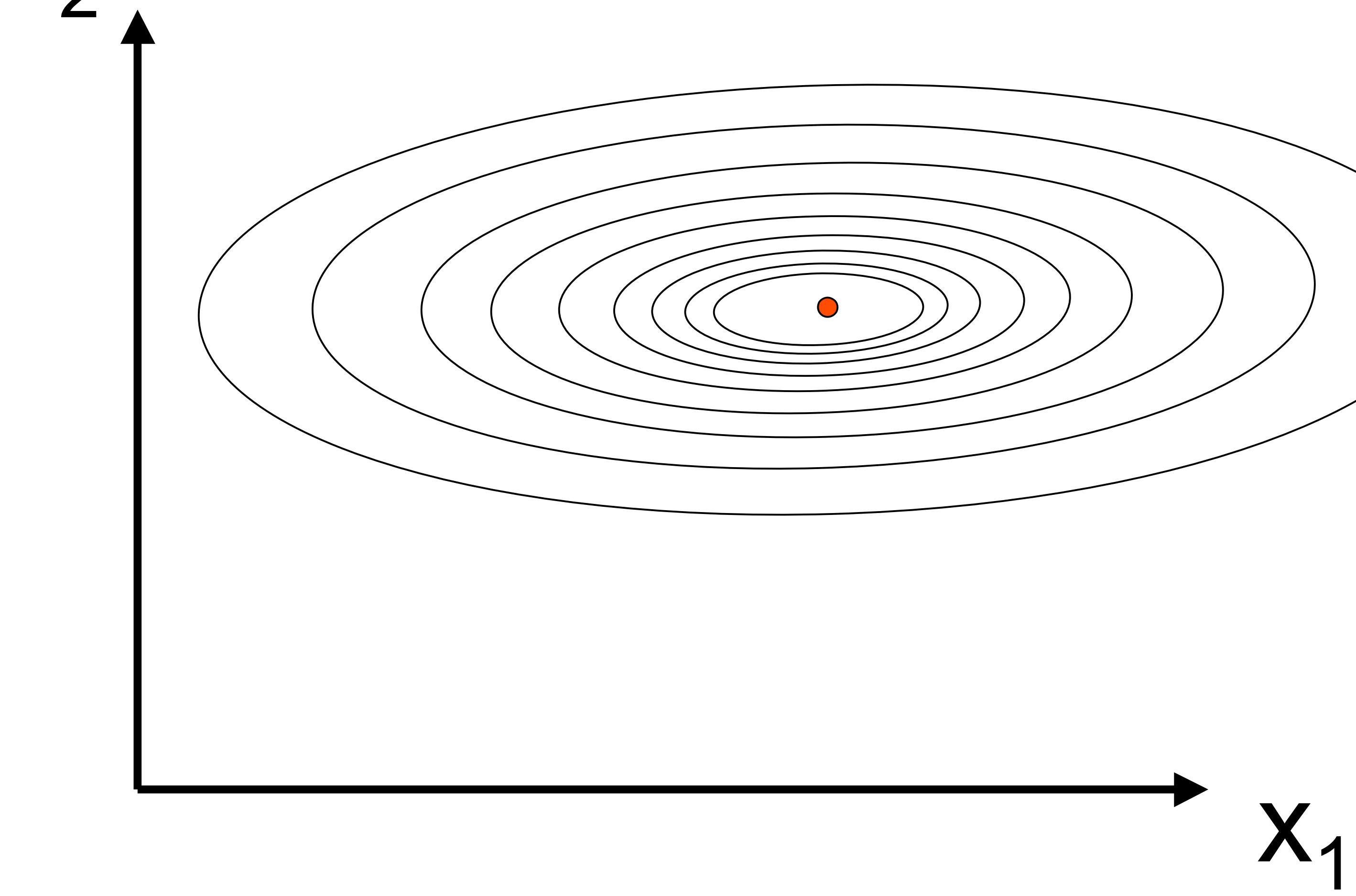
Choice of State Variables and Preconditioning

- 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

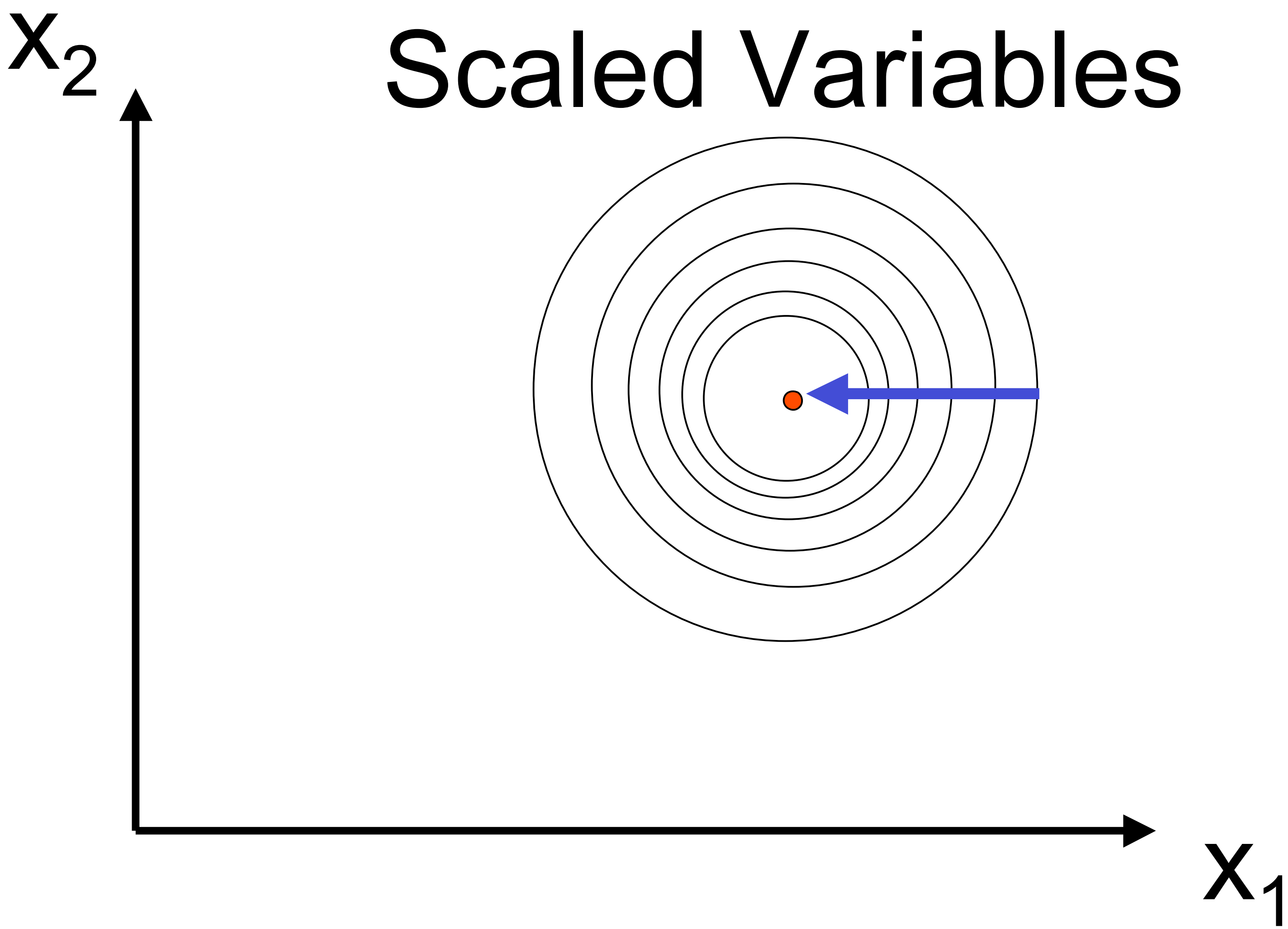
Cost Function for Correlated Errors



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.

$$\mathbf{B} \longrightarrow \mathbf{B(t)}$$

Analysis step as in OI

Evolution of Covariance Matrices

$$\mathbf{x}_b^{n+1} = M(\mathbf{x}_a^{n}) \quad \text{where } M \text{ is the non-linear model}$$

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

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

$$\text{Subtract: } \quad \varepsilon_b^{n+1} = \mathbf{M}\varepsilon_a^{n} + \varepsilon_m$$

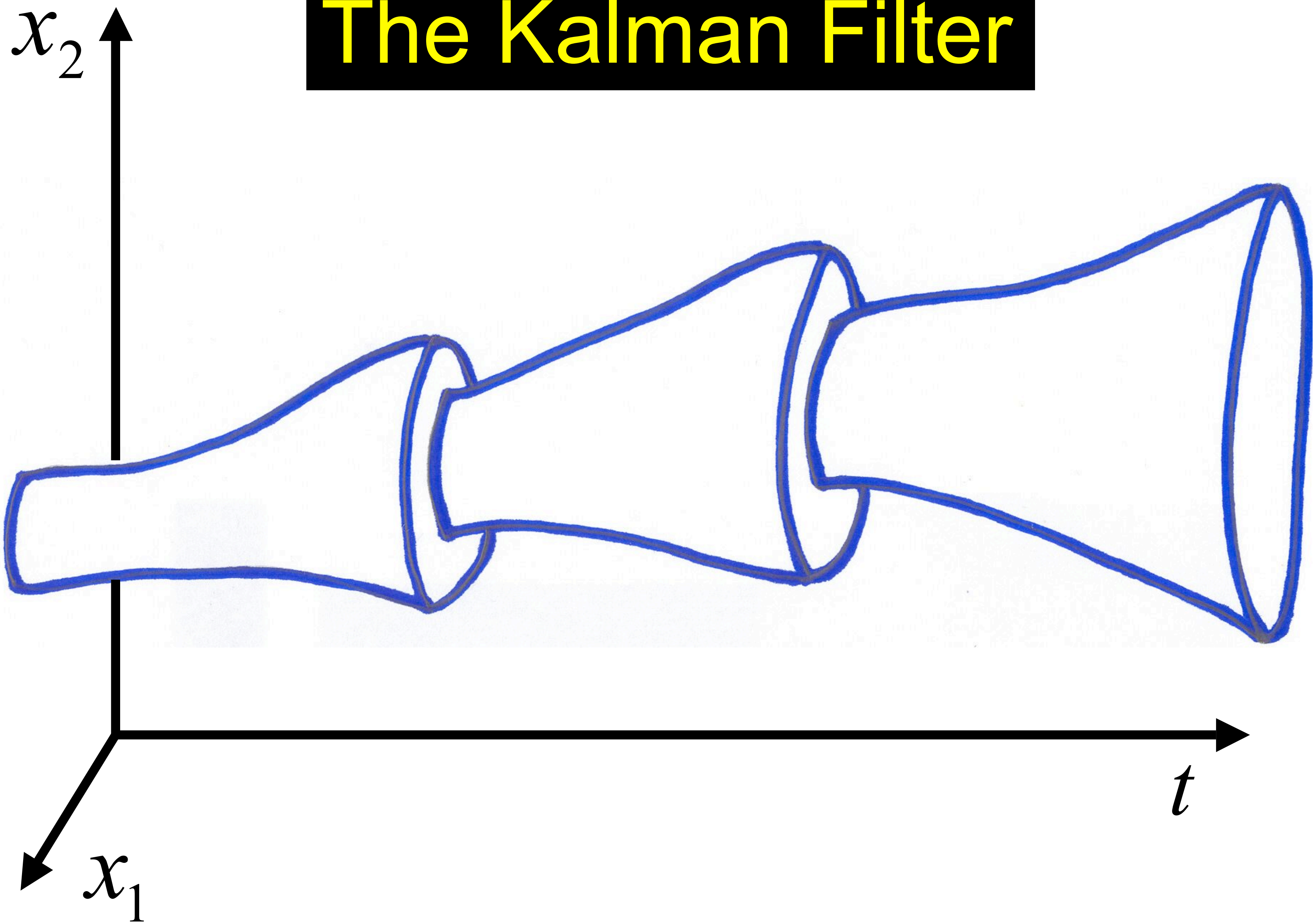
$$\text{The forecast error covariance is: } \mathbf{B} = \mathbf{P}_f(t_{n+1})$$

$$= \langle (\varepsilon_b^{n+1})(\varepsilon_b^{n+1})^{\mathrm{T}} \rangle$$

$$= \mathbf{M}(t_n)\mathbf{P}_a\mathbf{M}^{\mathrm{T}}(t_n) + \mathbf{Q}(t_n) \quad \text{where } \mathbf{Q} = \langle \varepsilon_m \varepsilon_m^{\mathrm{T}} \rangle$$

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

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

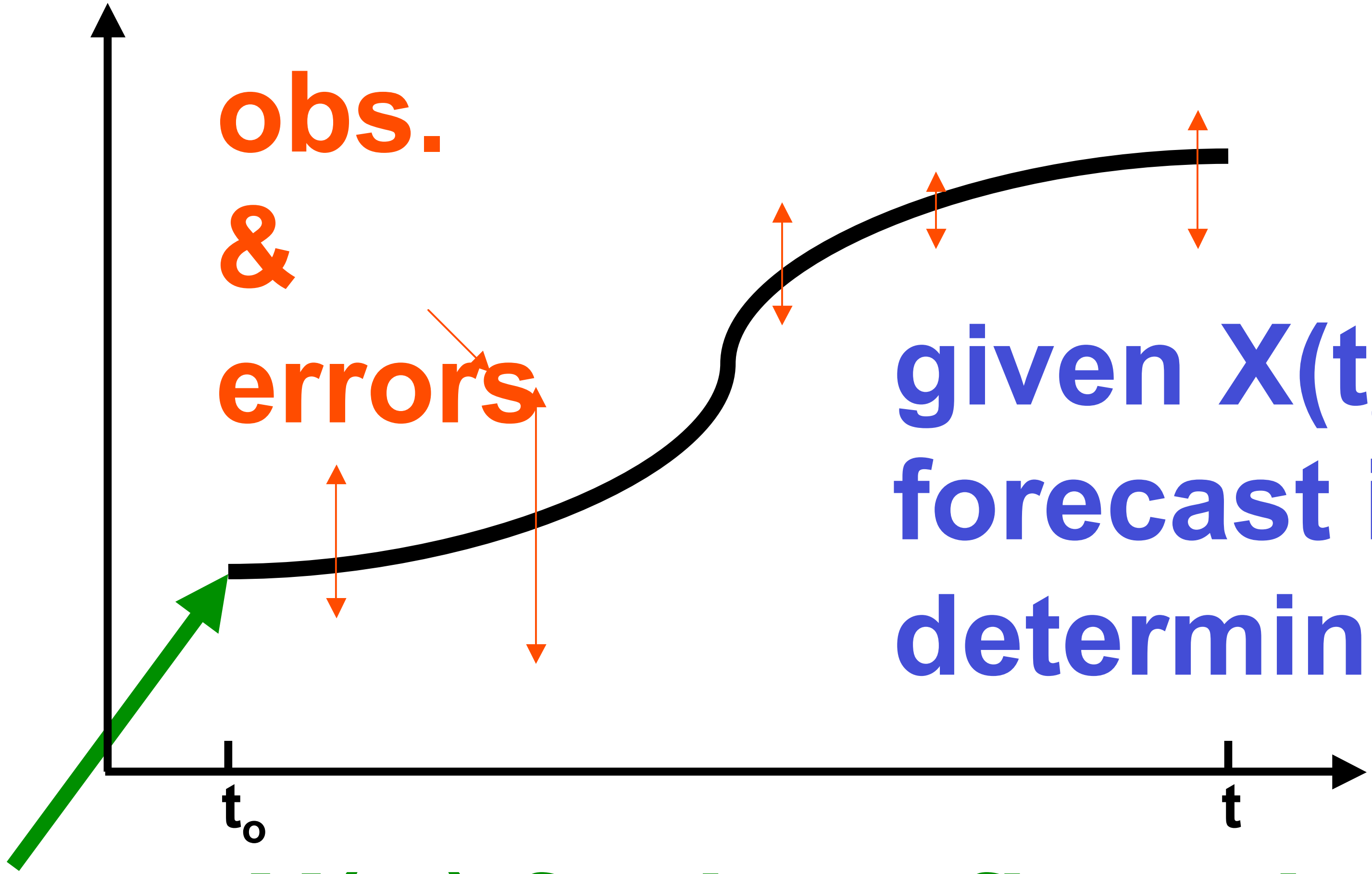
- Carry forecast error covariance matrix forward in time by using ensembles of forecasts:

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

- 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

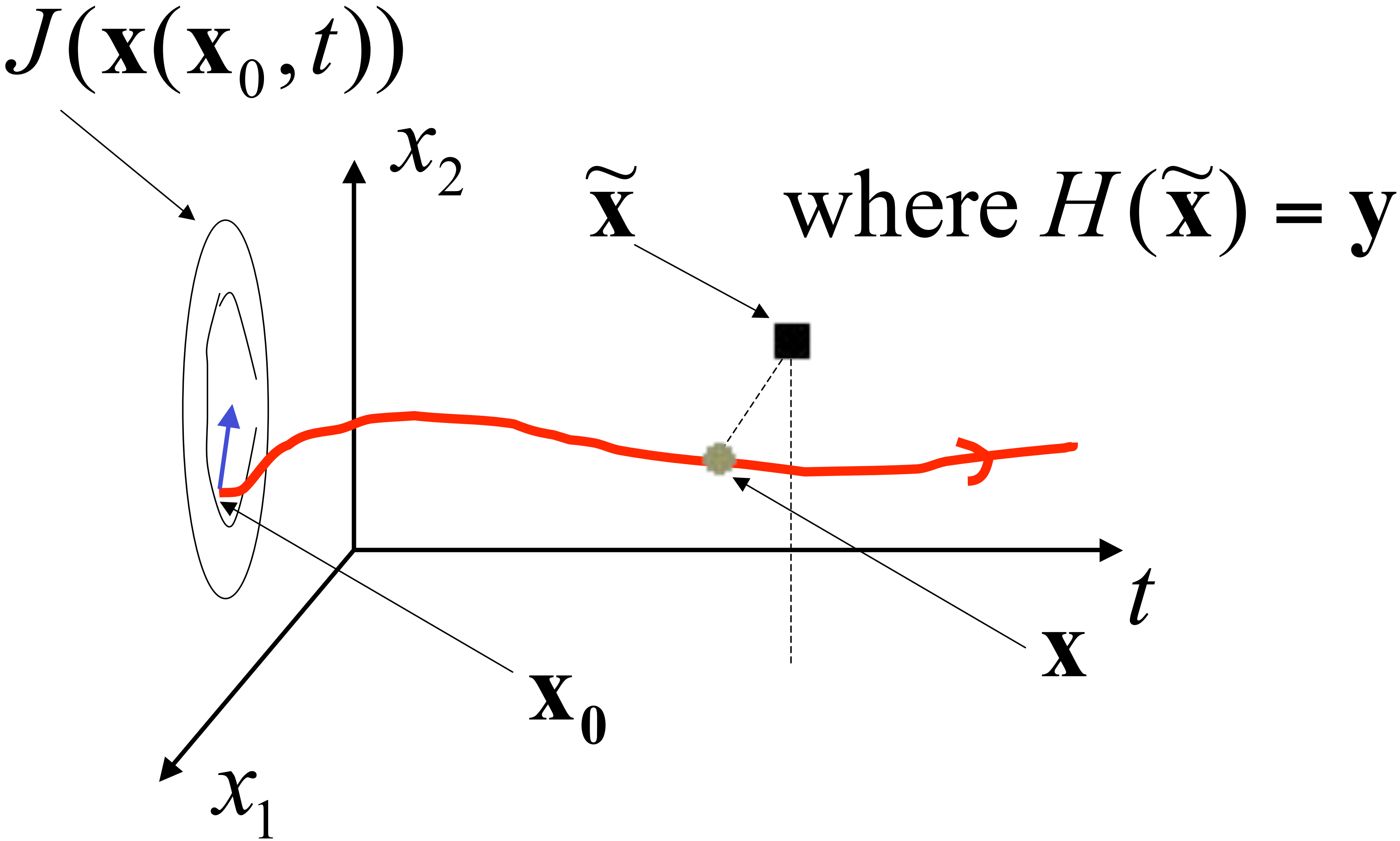
4D Variational Data Assimilation



given $X(t_o)$, the forecast is deterministic

vary $X(t_o)$ for best fit to data

4d-VAR For Single Observation at time *t*



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.

Some Matrix Algebra

adjoint of the model

$M : \mathbf{x}_0 \mapsto \mathbf{x}$

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

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

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

adjoint of tangent 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

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

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

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

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