Met Office

A Bayesian view of Data Assimilation

JCSDA Summer School.

Andrew Lorenc, Stevenson WA. July 2009.

Thanks to Jim Purser who introduced me to the Bayesian approach.

Content

1. Bayes Theorem – adding information
   - Gaussian PDFs
2. Simplest possible Bayesian NWP analysis
   - Two gridpoints, one observation.
3. Issues in practical implementation
   - Modelling/representing prior background error covariances **B**
   - Solving the large matrix equation
   - Estimating **B**.
4. Predicting the prior PDF
   - a Bayesian view of 4D-Var v Ensemble KF

Bayes Theorem – adding information

Gaussian PDFs

(Non-Gaussian observational errors - Quality Control will be covered in another lecture.)

Bayes' Theorem for Discrete Events

$A \quad B$	events
$P(A)$	probability of A occurring, or knowledge about A's past occurrence
$P(A \cap B)$	probability that A and B both occur,
$P(A \mid B)$	conditional probability of A given B

We have two ways of expressing $P(A \cap B)$:

$$P(A \cap B) = P(B)\,P(A \mid B) = P(A)\,P(B \mid A)$$

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

Can calculate $P(B)$ from: $$P(B) = P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})$$

Bayes theorem in continuous form,

to estimate a value x given an observation y^o

$$p(x \mid y^o) = \frac{p(y^o \mid x)p(x)}{p(y^o)}$$

$p(x \mid y^o)$ is the posterior distribution,

$p(x)$ is the prior distribution,

$p(y^o \mid x)$ is the likelihood function for *x*

Can get $p(y^o)$ by integrating over all x: $p(y^o) = \int p(y^o \mid x)p(x)dx$

Assume Gaussian pdfs

Prior is Gaussian with mean x^b, variance V_b : $\quad x \sim N(x^b, V_b)$

$$p(x) = (2\pi V_b)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{(x - x^b)^2}{V_b} \right)$$

Ob y^o, Gaussian about true value x variance V_o : $\quad y^o \sim N(x, V_o)$

$$p(y^o \mid x) = (2\pi V_o)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{(y^o - x)^2}{V_o} \right)$$

Substituting gives a Gaussian posterior: $\quad x \sim N(x^a, V_a)$

$$p(x) = (2\pi V_a)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{(x - x^a)^2}{V_a} \right)$$

Advantages of Gaussian assumption

1. Best estimate is a found by solving linear equations:

$$\frac{1}{V_a} = \frac{1}{V_o} + \frac{1}{V_b} \qquad\qquad \frac{1}{V_a} x^a = \frac{1}{V_o} y^o + \frac{1}{V_b} x^b$$

$$p(x) = \left(2\pi V_a\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{(x - x^a)^2}{V_a} \right)$$

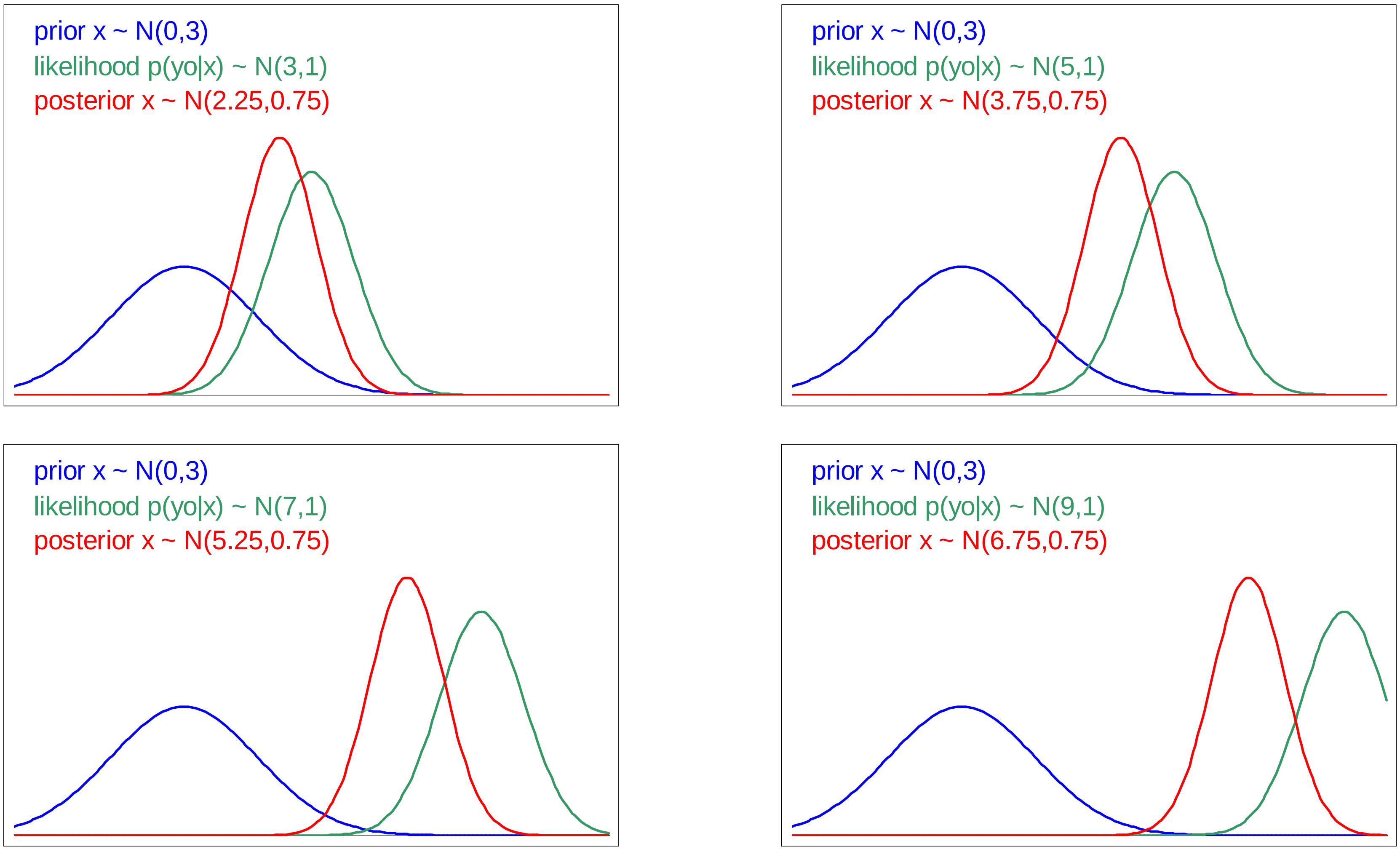
Taking logs gives quadratic equation; differentiating to find extremum gives linear equation.

2. Best estimate is a function of values & [co-]variances only.

Often these are all we know.

3. Weights are independent of values.

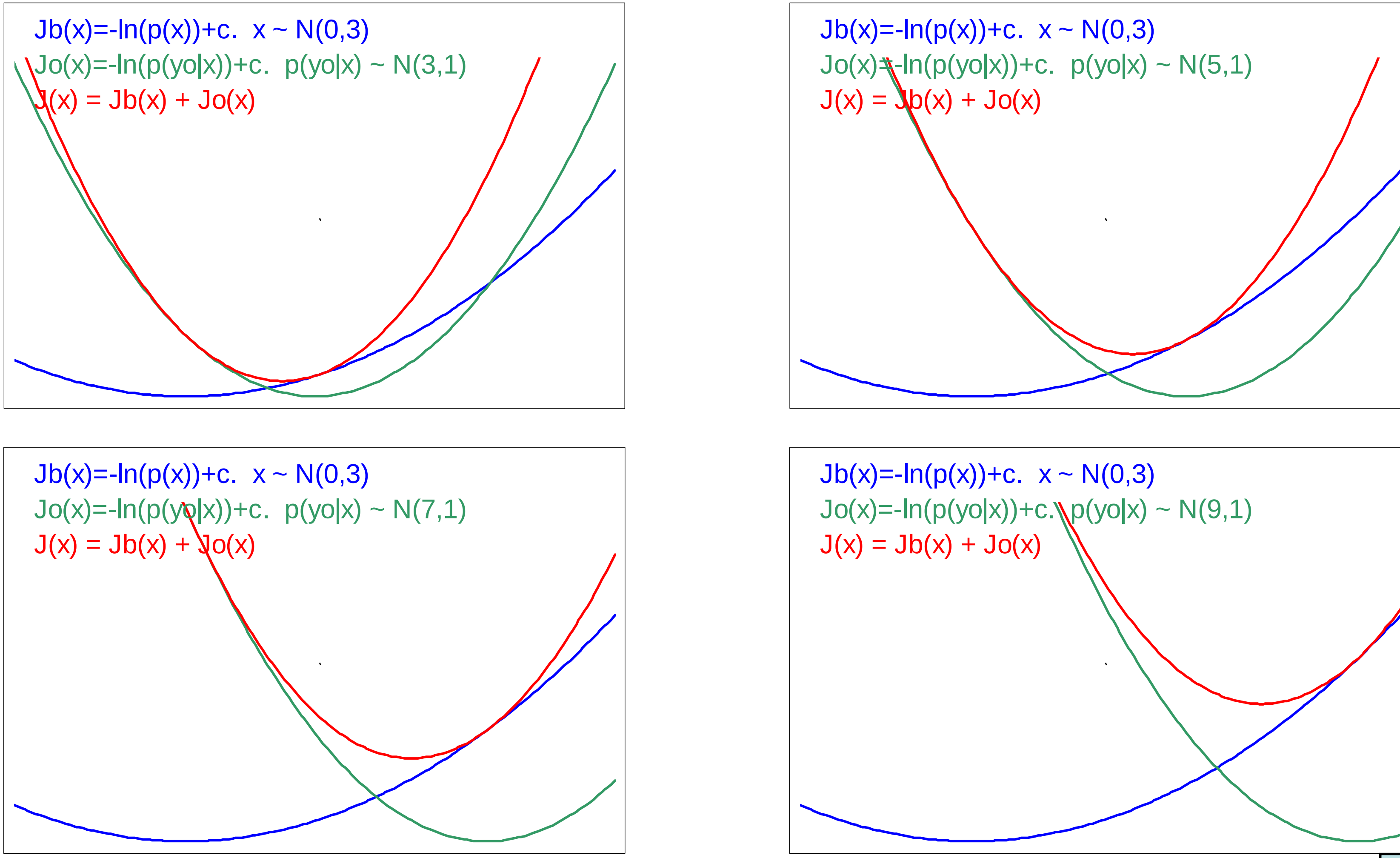
Combination of Gaussian prior & observation
- Gaussian posterior,
- weights independent of values.



Variational Penalty Functions

- Finding the most probable posterior value involves maximising a product [of Gaussians]
- By taking $-ln$ of the posterior PDF, we can instead minimise a sum [of quadratics]
- This is often called the "Penalty Function" J
- Additive constants can be ignored

Penalty functions: $J(x) = -\ln(p(x))+c$
p Gaussian $\Rightarrow J$ quadratic



PDFs

Simplest possible Bayesian NWP analysis

Simplest possible example – 2 grid-points, 1 observation. Standard notation:

Ide, K., Courtier, P., Ghil, M., and Lorenc, A.C. 1997: "Unified notation for data assimilation: Operational, Sequential and Variational" *J. Met. Soc. Japan*, Special issue "Data Assimilation in Meteorology and Oceanography: Theory and Practice." **75**, No. 1B, 181—189

Model is two grid points:
$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

1 observed value y^o midway (but use notation for >1):
$$\mathbf{y}^o = \begin{pmatrix} y^o \end{pmatrix}$$

Can interpolate an estimate $\mathbf{y}$ of the observed value:

$$\mathbf{y} = H(\mathbf{x}) = \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 = \mathbf{H}\mathbf{x} = \begin{pmatrix} \tfrac{1}{2} & \tfrac{1}{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

This example H is linear, so we can use matrix notation for fields as well as increments.

background pdf

We have prior estimate x^b_1 with error variance V_b:

$$p(x_1) = (2\pi V_b)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}\left(x_1 - x_1^b\right)^2 / V_b\right)$$

$$p(x_2) = (2\pi V_b)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}\left(x_2 - x_2^b\right)^2 / V_b\right)$$

But errors in x_1 and x_2 are usually correlated
$\Rightarrow$ must use a multi-dimensional Gaussian:

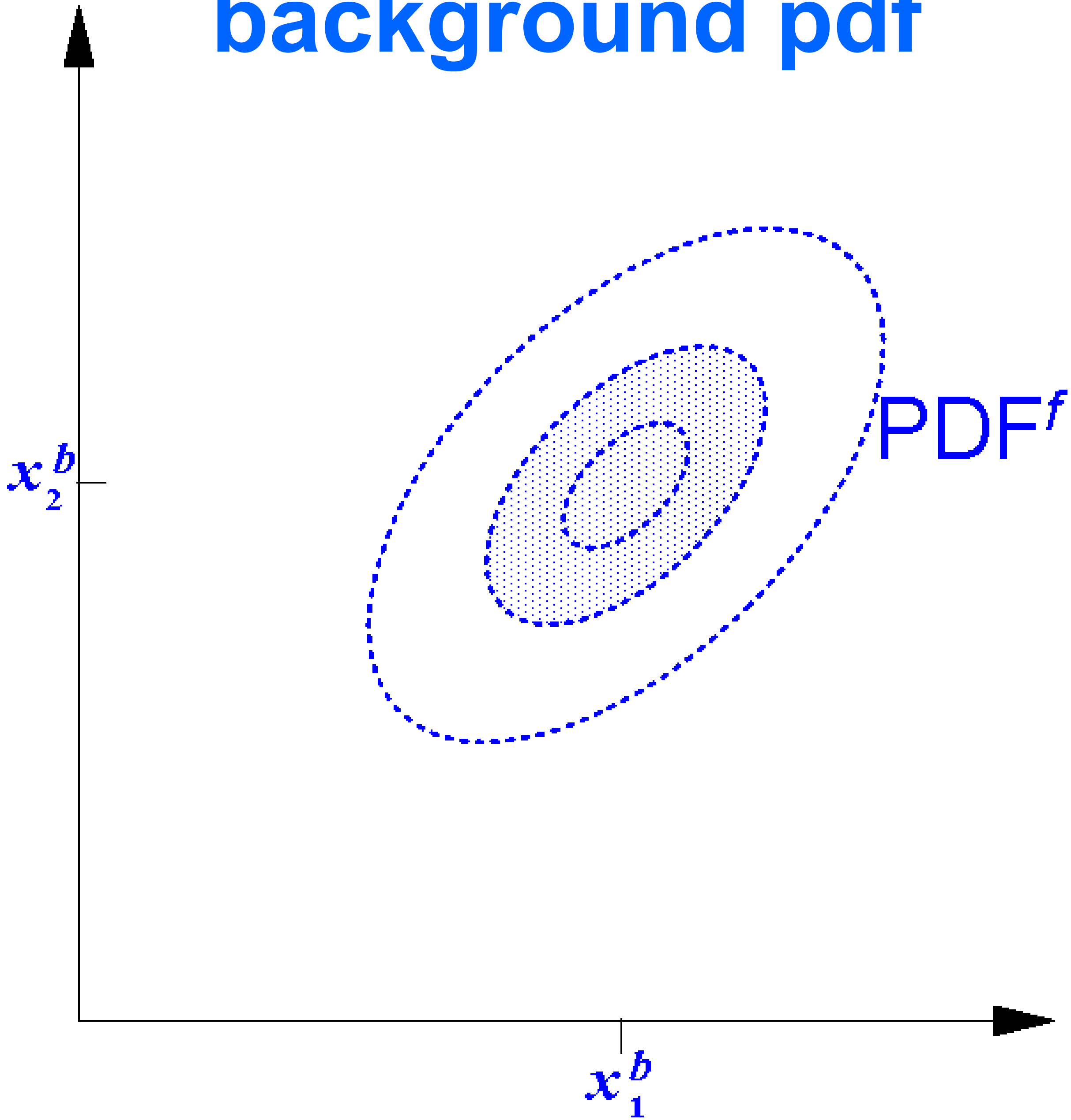
$$\mathbf{x} \sim N(\mathbf{x} : \mathbf{x}^b, \mathbf{B})$$

$$p(x_1 \cap x_2) = p(\mathbf{x}) = \left((2\pi)^2 |\mathbf{B}|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{x} - \mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^b\right)\right)$$

where $\mathbf{B}$ is the covariance matrix:

$$\mathbf{B} = V_b \begin{pmatrix} 1 & \mu \\ \mu & 1 \end{pmatrix}$$

background pdf



Observational errors

Lorenc, A.C. 1986: "Analysis methods for numerical weather prediction." *Quart. J. Roy. Met. Soc.*, **112**, 1177-1194.

instrumental error

$$\mathbf{y}^o \sim N\left(\mathbf{y}^t, \mathbf{E}\right)$$

$$p\left(\mathbf{y}^o \mid \mathbf{y}\right) = \left(2\pi\,|\mathbf{E}|\right)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}\left(\mathbf{y}^o - \mathbf{y}\right)^T \mathbf{E}^{-1}\left(\mathbf{y}^o - \mathbf{y}\right)\right)$$

error of representativeness

$$\mathbf{y} \sim N\left(H\left(\mathbf{x}^t\right), \mathbf{F}\right)$$

$$p_t\left(\mathbf{y} \mid \mathbf{x}^t\right) = \left(2\pi|\mathbf{F}|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{y} - H\left(\mathbf{x}^t\right)\right)^T \mathbf{F}^{-1}\left(\mathbf{y} - H\left(\mathbf{x}^t\right)\right)\right)$$

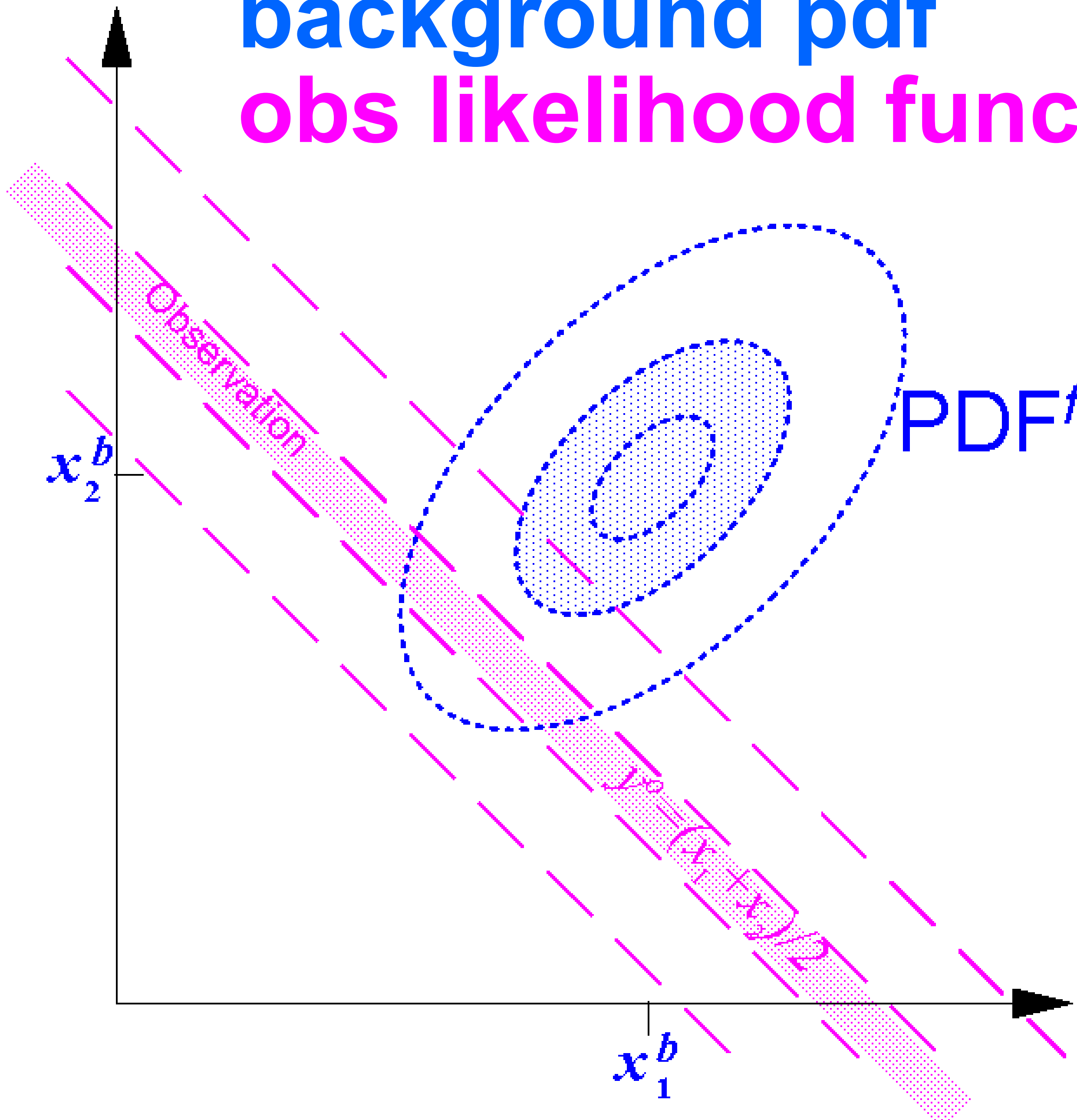
Observational error combines these 2 :

$$\mathbf{y}^o \sim N\left(H\left(\mathbf{x}^t\right), \mathbf{E}+\mathbf{F}\right)$$

$$p\left(\mathbf{y}^o \mid \mathbf{x}^t\right) = \int p\left(\mathbf{y}^o \mid \mathbf{y}\right) p_t\left(\mathbf{y} \mid \mathbf{x}^t\right) \mathbf{dy}$$

$$= \left(2\pi|\mathbf{E}+\mathbf{F}|\right)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}\left(\mathbf{y}^o - H\left(\mathbf{x}^t\right)\right)^T \left(\mathbf{E}+\mathbf{F}\right)^{-1}\left(\mathbf{y}^o - H\left(\mathbf{x}^t\right)\right)\right)$$

background pdf
obs likelihood function



Bayesian analysis equation

$$p\left(\mathbf{x}|\mathbf{y}^o\right) = \frac{p\left(\mathbf{y}^o|\mathbf{x}\right)p\left(\mathbf{x}\right)}{p\left(\mathbf{y}^o\right)}$$

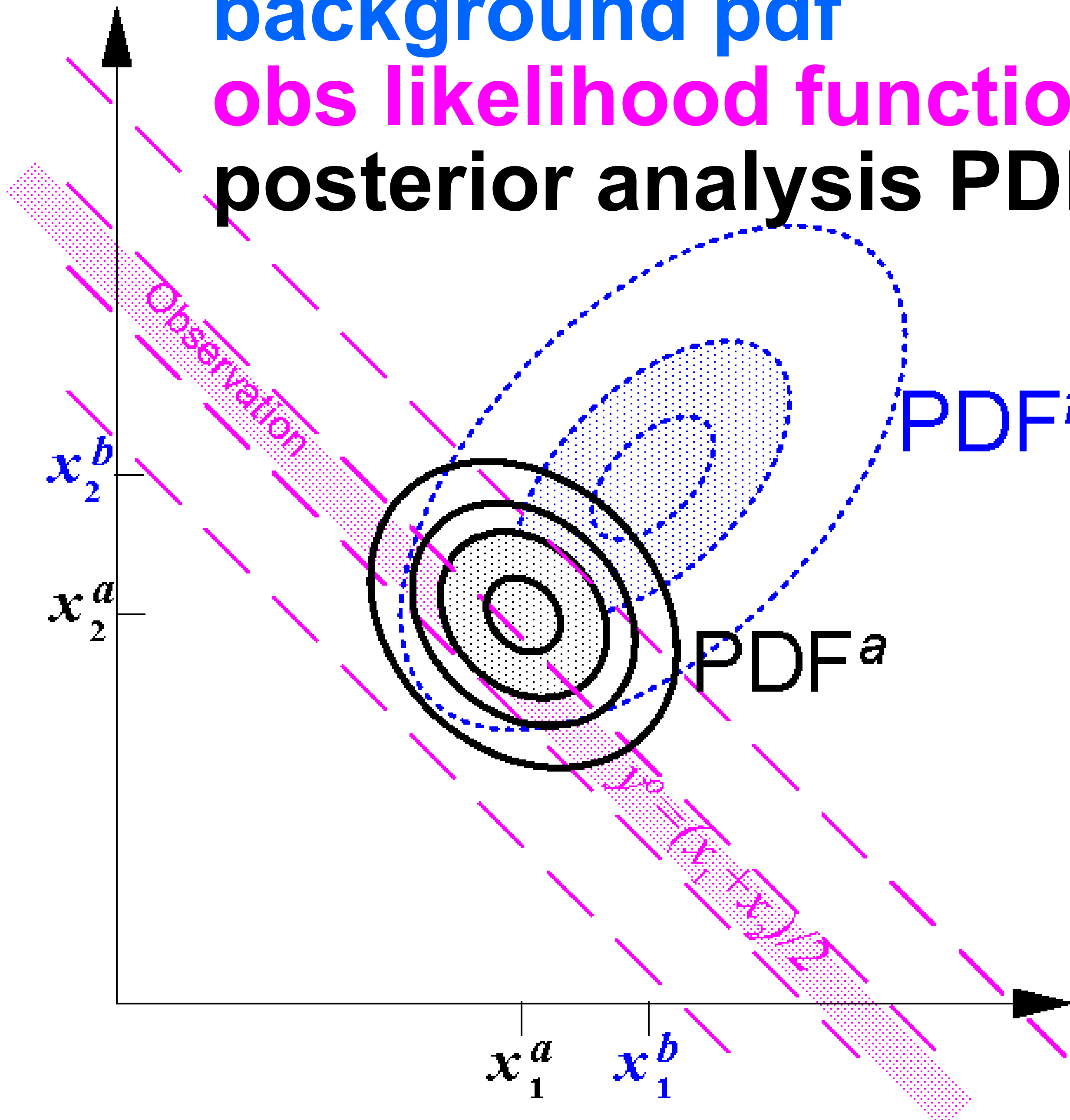
Property of Gaussians that, if H is linearisable : $\mathbf{x} \sim N\left(\mathbf{x}^a, \mathbf{A}\right)$

where x^a and $\mathbf{A}$ are defined by:

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\left(\mathbf{E}+\mathbf{F}\right)^{-1}\mathbf{H}$$

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A}\mathbf{H}^T\left(\mathbf{E}+\mathbf{F}\right)^{-1}\left(\mathbf{y}^o - H(\mathbf{x}^b)\right)$$

background pdf
obs likelihood function
posterior analysis PDF



Analysis equation

For our simple example the algebra is easily done by hand, without manipulating matrices, giving:

$$\mathbf{x}^a = \begin{pmatrix} x_1^a \\ \\ x_2^a \end{pmatrix} = \begin{pmatrix} x_1^b \\ \\ x_2^b \end{pmatrix} + \frac{\left(V^b\left(\frac{1+\mu}{2}\right)\right)^2}{\mathbf{E}+\mathbf{F}+V^b\left(\frac{1+\mu}{2}\right)}\left[y^o - \frac{x_1^b + x_2^b}{2}\right]\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Practical implementation of the Bayesian Analysis Equation

Issues in practical implementation
"*The devil is in the details*"

There are significantly different choices possible for each of the following options. The combinations of these choices make up a very wide range and large number of analysis schemes, all implementing the same Bayesian equation!

- Modelling and representing prior background error covariances **B.**
- Expressing the equations in a form amenable to solution.
- Computing the solution.
- Estimating **B**.

Michael Ghil on OI & Kalman Filter

Reprinted from Preprint Volume: Fifth Conference on Numerical Weather Prediction, Nov. 2-6, 1981; Monterey, Calif. Published by the American Meteorological Society, Boston, Mass.

OPTIMAL INTERPOLATION AND THE KALMAN FILTER are similar,

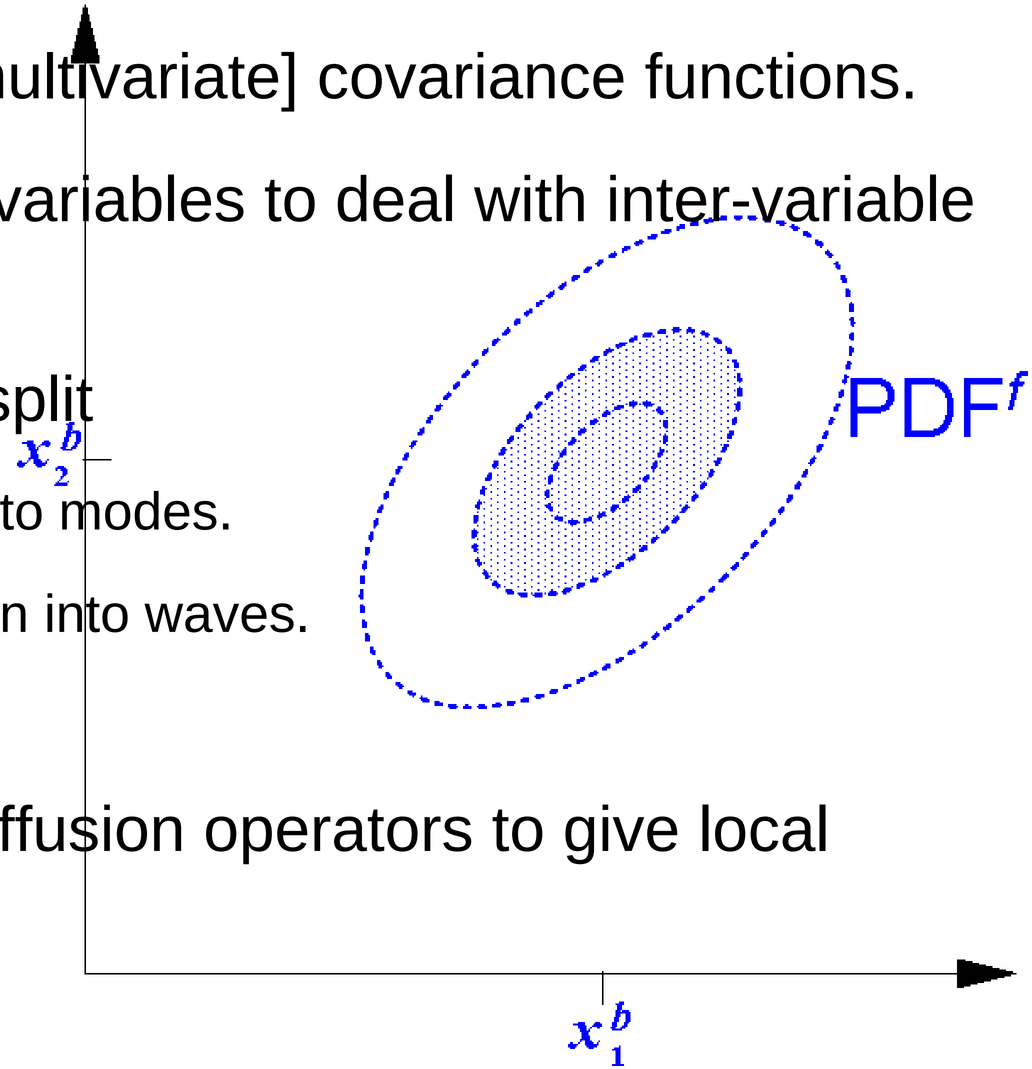
but the fun is in the difference.

S. Cohn, M. Ghil* and E. Isaacson

Courant Institute of Mathematical Sciences
New York University, New York, New York 10012

Modelling and representing prior background error covariances **B.**

- Explicit point-point [multivariate] covariance functions.
- Transformed control variables to deal with inter-variable covariances.
- Vertical – horizontal split
  - EOF decomposition into modes.
  - Spectral decomposition into waves.
  - Wavelets.
- Recursive filters or diffusion operators to give local variations.
- Ensemble members.



Schlatter's (1975) multivariate

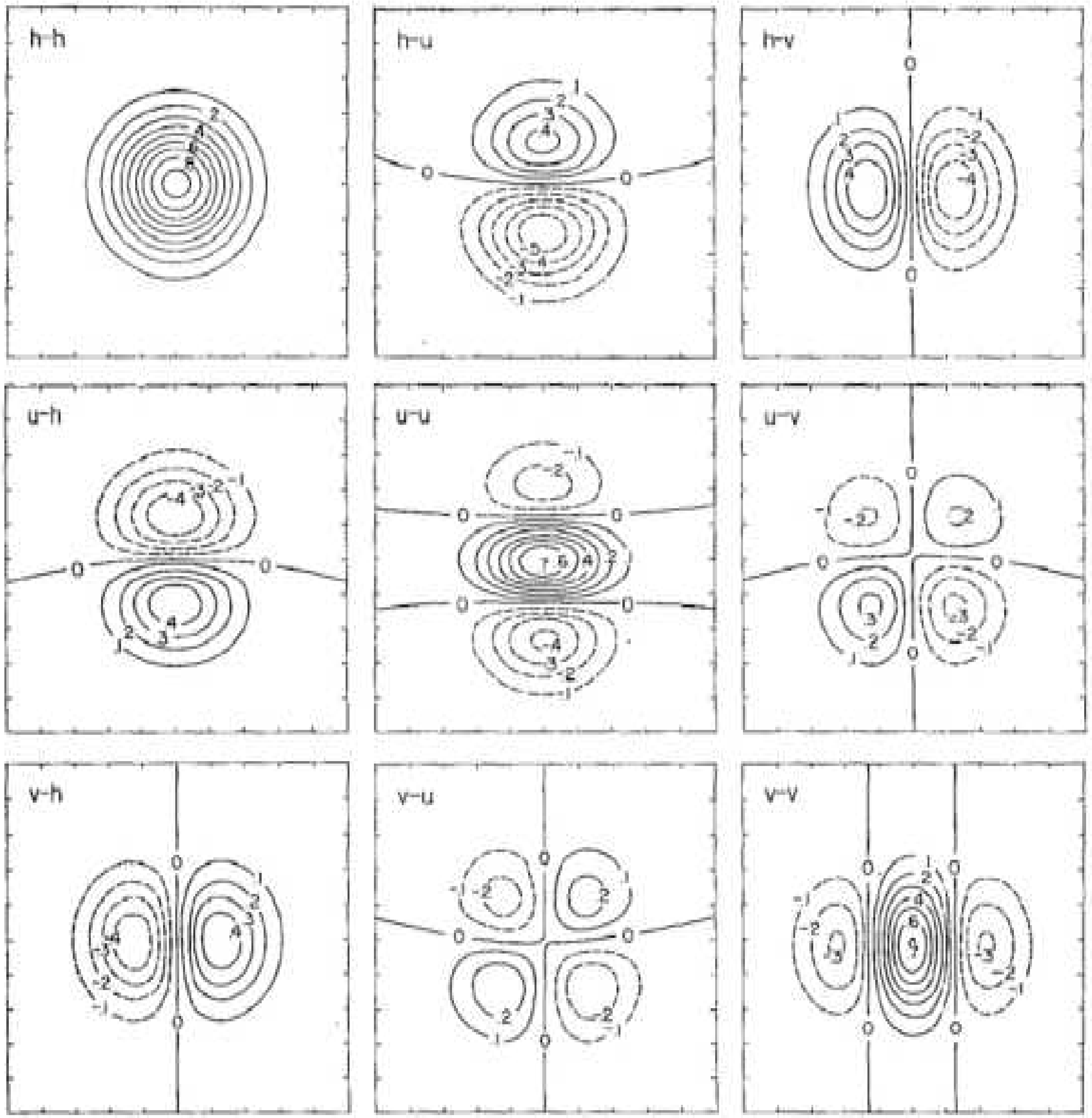


FIG. 3. Correlations among the variables h, u, and v based upon the expression $\mu = 0.95\exp(-1.24s^2)$ for height-height correlation and the geostrophic relations. Diagrams centered at 110°W, 35°N. Tick marks 500 km apart.

Transformed control variable.

- Look for a "balanced" variable from which we can calculate balanced flow in all variables: streamfunction, PV.

$$\delta\mathbf{x} = \mathbf{U}_p \begin{pmatrix} \psi \\ \chi \\ \mathbf{b} \end{pmatrix} = \mathbf{U}_p \mathbf{v}$$

- Define transforms from (**U**) or to (**T**) this variable and a residual variable, which by construction/hypothesis is uncorrelated making **B** block diagonal. (Compare EOFs)

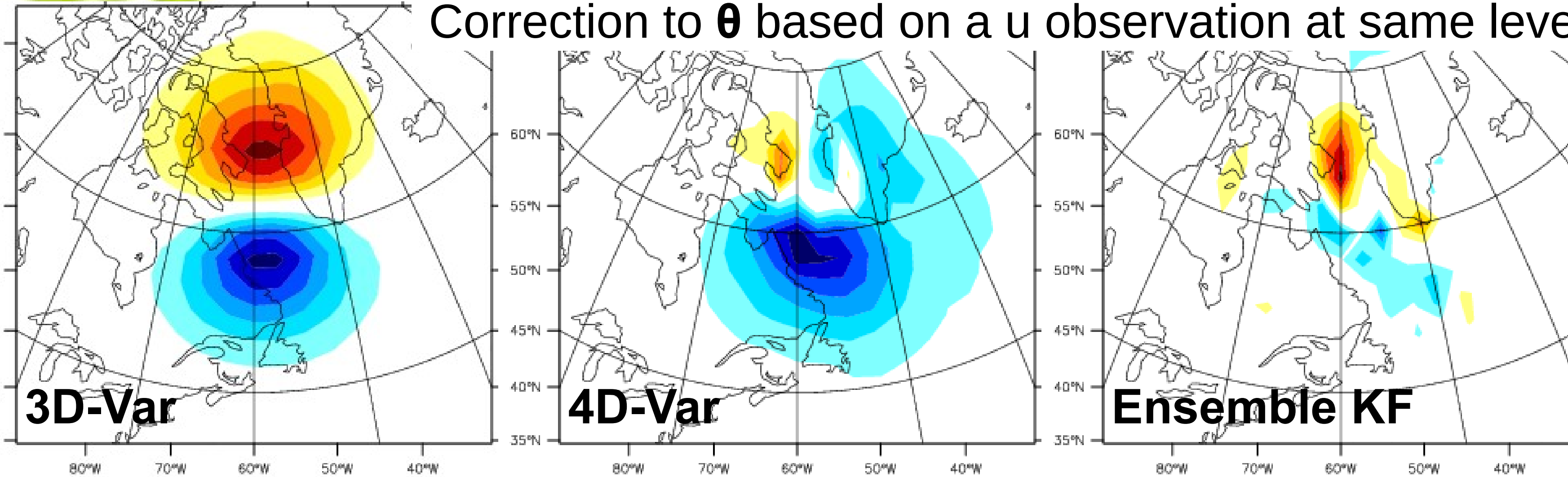
$$\mathbf{B}_{(\mathbf{v})} = \begin{pmatrix} \mathbf{B}_{(\psi)} & 0 & 0 \\ 0 & \mathbf{B}_{(\chi)} & 0 \\ 0 & 0 & \mathbf{B}_{(\mathbf{b})} \end{pmatrix}$$

$$\mathbf{B}_{(\mathbf{x})} = \mathbf{U}_p \mathbf{B}_{(\mathbf{v})} \mathbf{U}_p^{\ T}$$

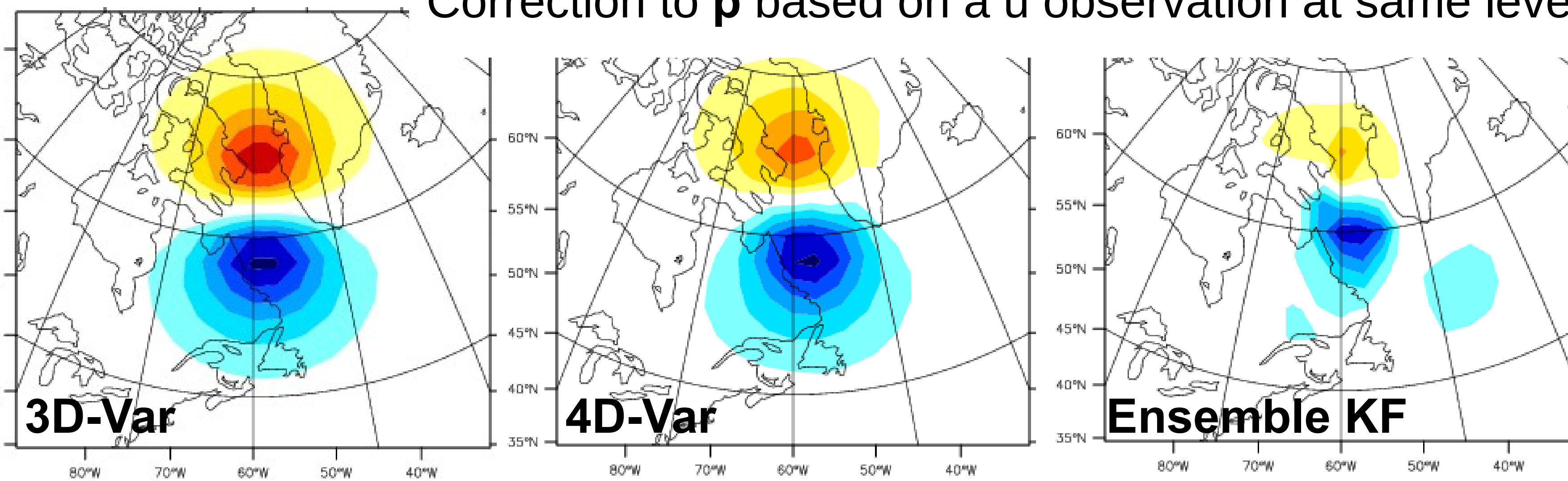
- Transformed variables still need spatial covariance model, but not multivariate. (Further transforms may be used to represent these.)

Comparison of covariance models

Correction to **θ** based on a u observation at same level



Correction to **p** based on a u observation at same level



Adam Clayton

Equations - all equivalent.

- Variational $\mathbf{x}^a \text{ minimises } J(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \tfrac{1}{2}(\mathbf{y}^o - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x}))$

$$\mathbf{A}^{-1} = \left(\frac{\partial^2 J}{\partial \mathbf{x}^2}\right) = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$$

- Kalman Filter. Kalman Gain=**K.** $\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\left(\mathbf{y}^o - H(\mathbf{x}^b)\right)$ $\qquad \mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}$

- Observation space $\mathbf{K} = \mathbf{BH}^T\left(\mathbf{HBH}^T + \mathbf{R}\right)^{-1}$

Demonstrate equivalence using Sherman–Morrison–Woodbury formula

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

- Ensemble space Square-root Filters, e.g. ETKF

$$\mathbf{B} = \mathbf{Z}^f\left(\mathbf{Z}^f\right)^T$$

$$\mathbf{Z}^a = \mathbf{Z}^f\mathbf{T}$$

$$\mathbf{A} = \mathbf{Z}^a\left(\mathbf{Z}^a\right)^T$$

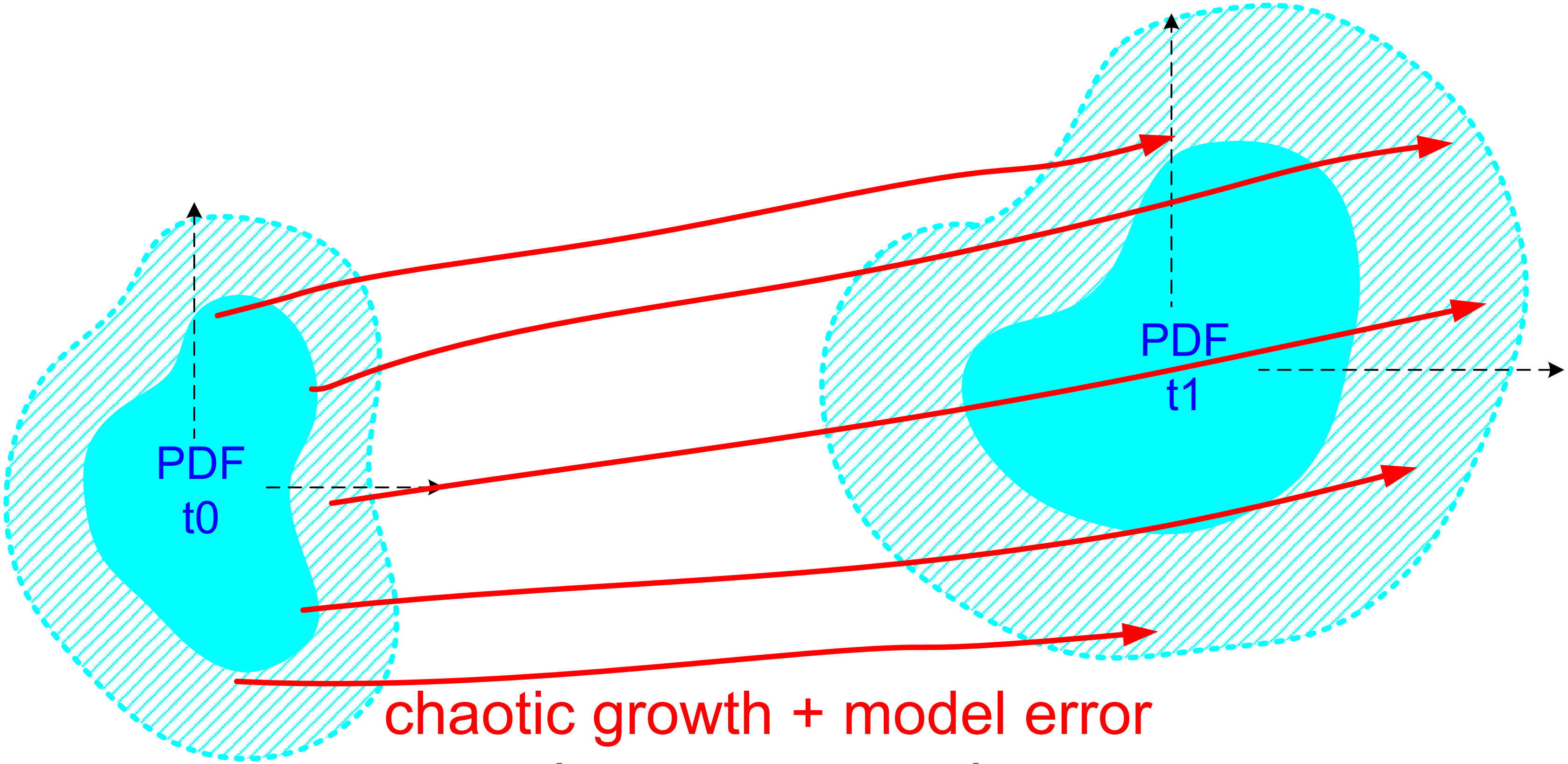
Estimating PDFs or covariances

- Even if we knew the "truth", we could never run enough experiments in the lifetime of an NWP system to estimate its error PDF, or even its error covariance $\mathbf{B}$.
- Simplifying assumptions are essential (e.g. Gaussian, ...)
- Even a simplified error model has so many parameters that we cannot determine them by NWP trials to determine which give the best forecasts.
- In practice we can only measure innovations – cannot get separate estimates of $\mathbf{B}$ & $\mathbf{R}$ without assumptions (Talagrand).
- ***Need to understand physics!***

How to estimate the prior PDF?
How to calculate its time evolution?

i.e. 4D-Var versus Ensemble KF

Fokker-Planck Equation



Ensemble methods attempt to sample entire PDF.

Gaussian Probability Distribution Functions

- Easier to fit to sampled errors.
- Quadratic optimisation problems, with linear solution methods – much more efficient.
- The Kalman filter is optimal for linear models, but
  - it is not affordable for expensive models (despite the “easy” quadratic problem)
  - it is not optimal for nonlinear models.
- Advanced methods based on the Kalman filter can be made affordable:
  - Ensemble Kalman filter (EnKF, ETKF, ...)
  - Four-dimensional variational assimilation (4D-Var)

Extended Kalman Filter

$\mathbf{x}$ is mean of PDF, $\mathbf{P}$ is covariance.

Analysis step

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{P}^f(t_i)\mathbf{H}_i^T\left(\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right)^{-1}\left(\mathbf{y}_i^o - H_i\left(\mathbf{x}^f(t_i)\right)\right)$$

$$\mathbf{P}^a(t_i) = \mathbf{P}^f(t_i) - \mathbf{P}^f(t_i)\mathbf{H}_i^T\left(\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right)^{-1}\mathbf{H}_i\mathbf{P}^f(t_i)$$

Forecast step

$$\mathbf{x}^f(t_{i+1}) = M_i\left(\mathbf{x}^a(t_i)\right)$$

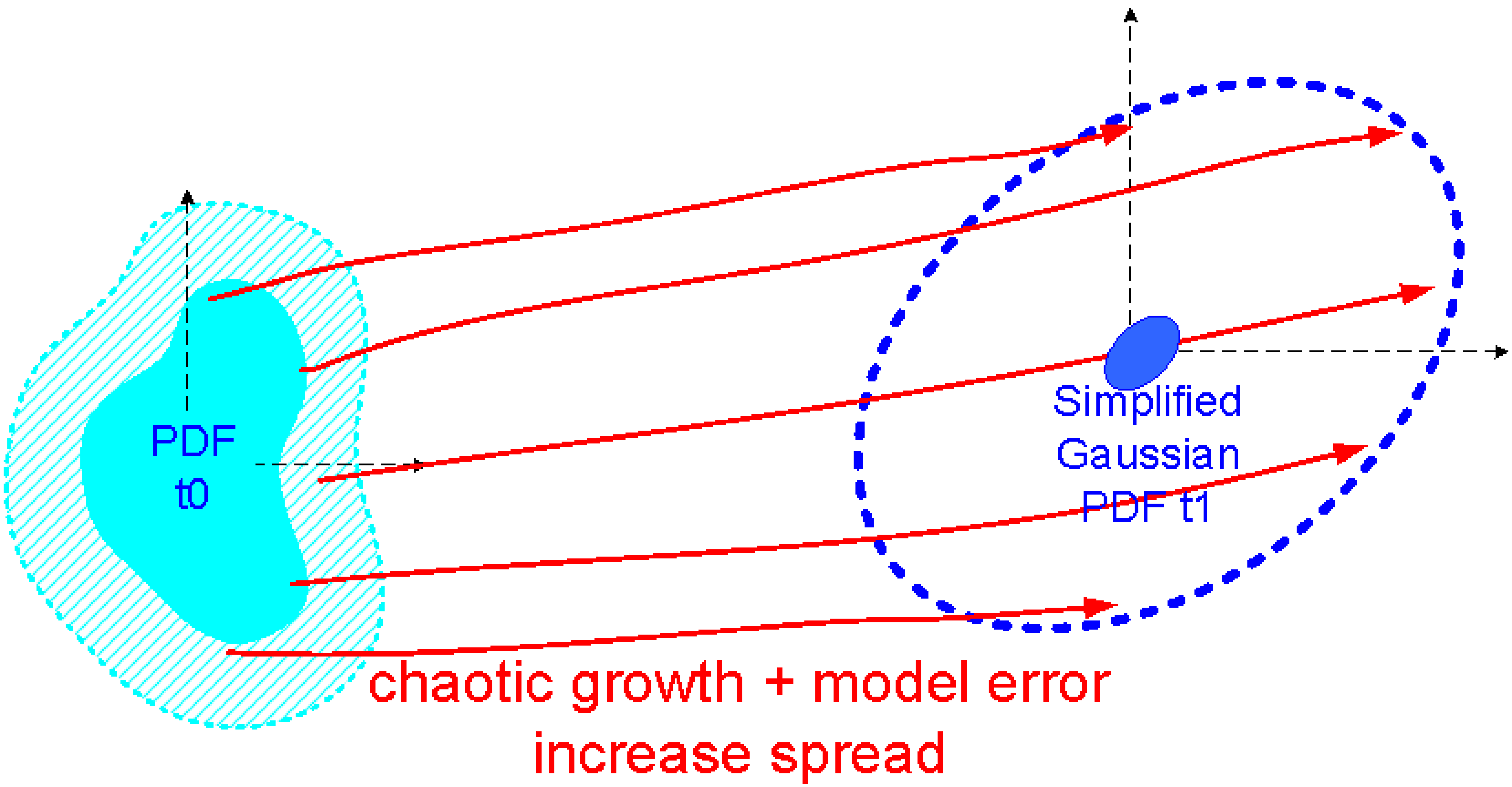
True discretised dynamics $\mathbf{x}^t$ assumed to differ by stochastic perturbations:

$$\mathbf{x}^t(t_{i+1}) = M_i\left(\mathbf{x}^t(t_i)\right) + \boldsymbol{\eta}(t_i)$$

where $\boldsymbol{\eta}$ is a noise process with zero mean and covariance matrix $\mathbf{Q}_i$.

$$\mathbf{P}^f(t_{i+1}) = \mathbf{M}_i\mathbf{P}^a(t_i)\mathbf{M}_i^T + \mathbf{Q}_i$$

Ensemble Kalman filter



Fit Gaussian to forecast ensemble.

The Ensemble Kalman Filter (EnKF)

Construct an ensemble $\{\mathbf{x}_i^f\},(i=1,\_\_\_,N)$:

$$\mathbf{P}^f = \mathbf{P}_e^f = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)^T},$$

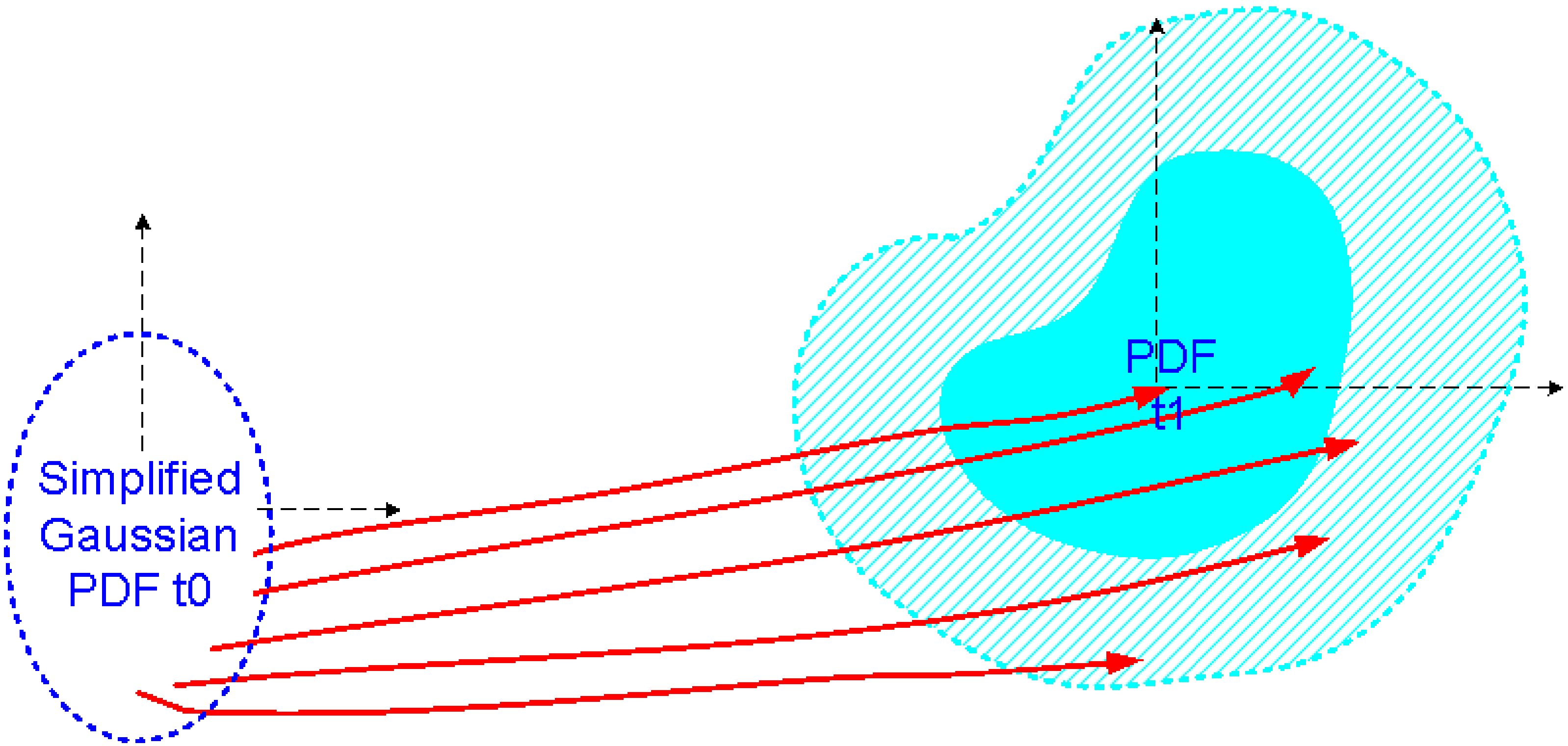
$$\mathbf{P}^f\mathbf{H}^T = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)^T},$$

$$\mathbf{H}\mathbf{P}^f\mathbf{H}^T = \overline{\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)^T}$$

Use these in the standard KF equation to update the best estimate (ensemble mean):

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - H\left(\overline{\mathbf{x}^f}\right)\right).$$

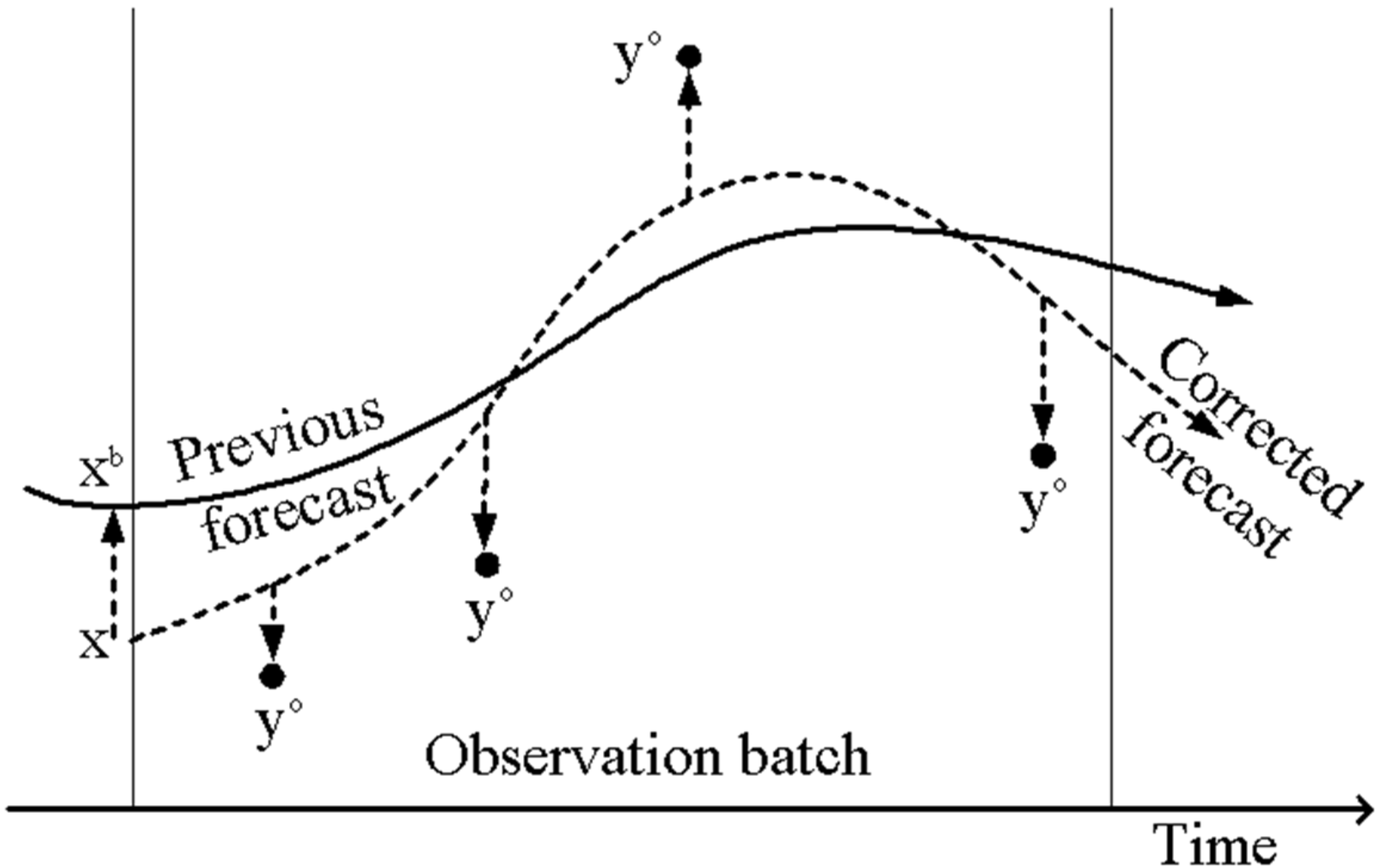
Deterministic 4D-Var



Initial PDF is approximated by a Gaussian.

Descent algorithm only explores a small part of the PDF, on the way to a local minimum.

Simple 4D-Var, as a least-squares best fit of a deterministic model trajectory to observations



Assumptions in deriving deterministic 4D-Var

Bayes Theorem - posterior PDF: $P\left(x \middle| y^o\right) = P\left(y^o \middle| x\right) P(x) \big/ P\left(y^o\right)$

where the obs likelihood function is given by: $P\left(y^o \middle| x\right) = f\left(y^o - y\right), \quad where \quad y = H(x)$

Impossible to evaluate the integrals necessary to find "best".

Instead assume best *x* maximises PDF, and minimises -ln(PDF): $J(x) = -\ln\left[P\left(y^o \middle| x\right)\right] - \ln\left[P(x)\right]$

Purser, R.J. 1984: "A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis". Preprints, 10th conference on weather forecasting and analysis. Am Met Soc. 102-105

*Lorenc, A.C. 1986: "Analysis methods for numerical weather prediction." Quart. J. Roy. Met. Soc., **112**, 1177-1194.*

The deterministic 4D-Var equations

$$P\left(\mathbf{x}\middle|\underline{\mathbf{y}}^o\right) \propto P(\mathbf{x})\,P\left(\underline{\mathbf{y}}^o\middle|\mathbf{x}\right)$$

Bayesian posterior pdf.

Assume Gaussians

$$P(\mathbf{x}) \propto \exp\left(-\tfrac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right)\right)$$

$$P\left(\underline{\mathbf{y}}^o\middle|\mathbf{x}\right) = P\left(\underline{\mathbf{y}}^o\middle|\underline{\mathbf{y}}\right) \propto \exp\left(-\tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)\right)$$

But nonlinear model makes pdf non-Gaussian: full pdf is too complicated to be allowed for.

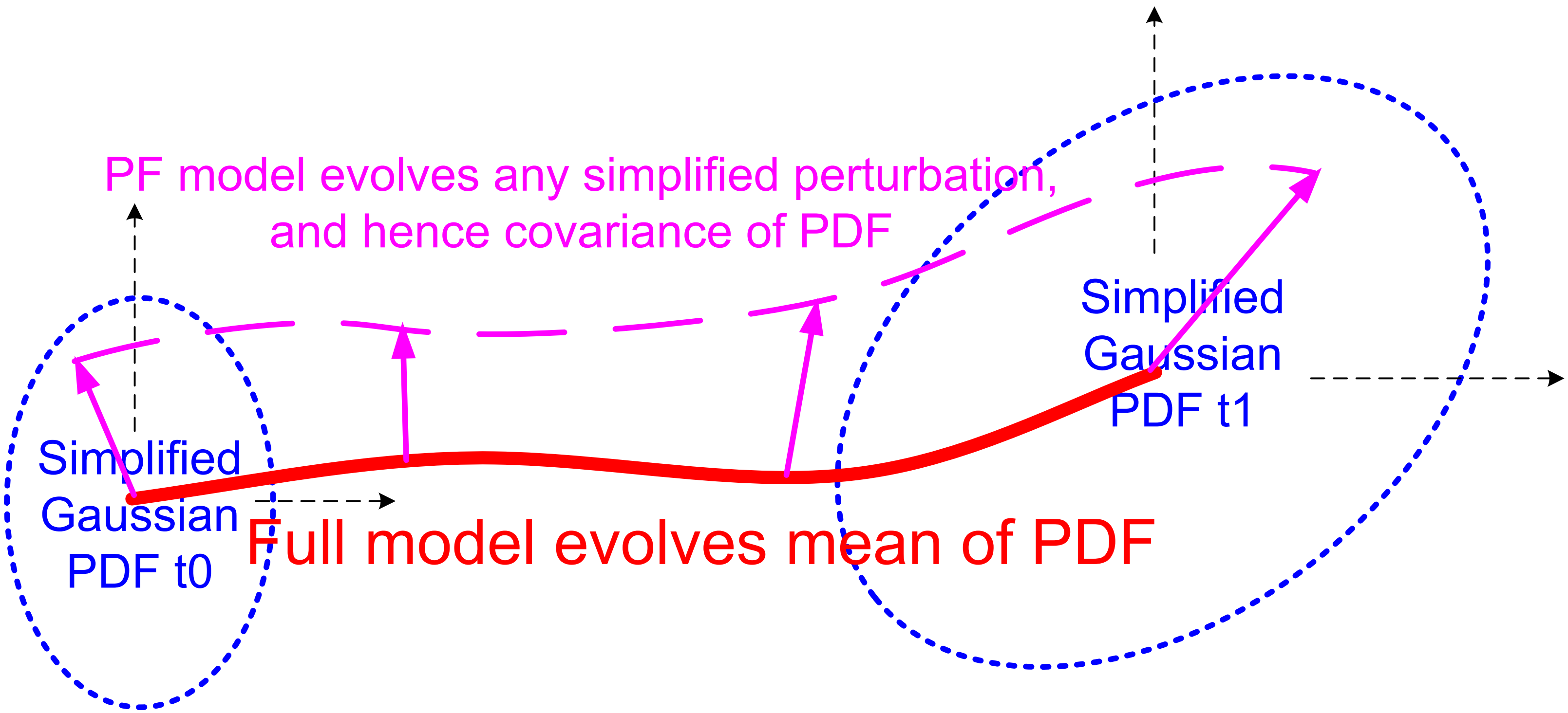
$$\underline{\mathbf{y}} = \underline{H}\left(\underline{M}(\mathbf{x})\right)$$

So seek mode of pdf by finding minimum of penalty function

$$J(\mathbf{x}) = \tfrac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) + \tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)$$

$$\nabla_{\mathbf{x}} J(\mathbf{x}) = \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) + \underline{\mathbf{M}}^*\underline{\mathbf{H}}^*\mathbf{R}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)$$

Statistical, incremental 4D-Var



Statistical 4D-Var approximates entire PDF by a Gaussian.

Statistical 4D-Var - equations

Independent, Gaussian background and model errors $\Rightarrow$ non-Gaussian pdf for general $\mathbf{y}$:

$$P\left(\delta\mathbf{x},\delta\underline{\boldsymbol{\eta}}\middle|\underline{\mathbf{y}}^o\right)\propto\exp\left(-\tfrac{1}{2}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)^T\mathbf{B}^{-1}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)\right)$$

$$\exp\left(-\tfrac{1}{2}\left(\delta\underline{\boldsymbol{\eta}}+\underline{\boldsymbol{\eta}}^g\right)^T\underline{\mathbf{Q}}^{-1}\left(\delta\underline{\boldsymbol{\eta}}+\underline{\boldsymbol{\eta}}^g\right)\right)$$

$$\exp\left(-\tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)\right)$$

Incremental linear approximations in forecasting model predictions of observed values converts this to an approximate Gaussian pdf:

$$\underline{\mathbf{y}}=\underline{\tilde{\mathbf{H}}}\underline{\tilde{\mathbf{M}}}\left(\delta\mathbf{x},\underline{\boldsymbol{\eta}}\right)+\underline{\bar{H}}\left(\underline{\bar{M}}\left(\mathbf{x}^g,\underline{\boldsymbol{\eta}}^g\right)\right)$$

The mean of this approximate pdf is identical to the mode, so it can be found by minimising:

$$J\left(\delta\mathbf{x},\delta\underline{\boldsymbol{\eta}}\right)=\tfrac{1}{2}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)^T\mathbf{B}^{-1}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)$$

$$+\tfrac{1}{2}\left(\delta\underline{\boldsymbol{\eta}}+\underline{\boldsymbol{\eta}}^g\right)^T\underline{\mathbf{Q}}^{-1}\left(\delta\underline{\boldsymbol{\eta}}+\underline{\boldsymbol{\eta}}^g\right)$$

$$+\tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)$$

Questions and answers