A Bayesian view of Data Assimilation and Quality Control of Observations

Lecture, Intensive Course on Data Assimilation

Andrew Lorenc, Buenos Aires, Argentina. October 27 - November 7, 2008

Content

1. Bayes Theorem – adding information
    - Gaussian PDFs
    - Non-Gaussian observational errors - Quality Control
2. Simplest possible Bayesian NWP analysis
    - Two gridpoints, one observation.
3. 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

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

Met Office

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

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

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

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

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

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

Advantages of Gaussian assumption

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

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

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

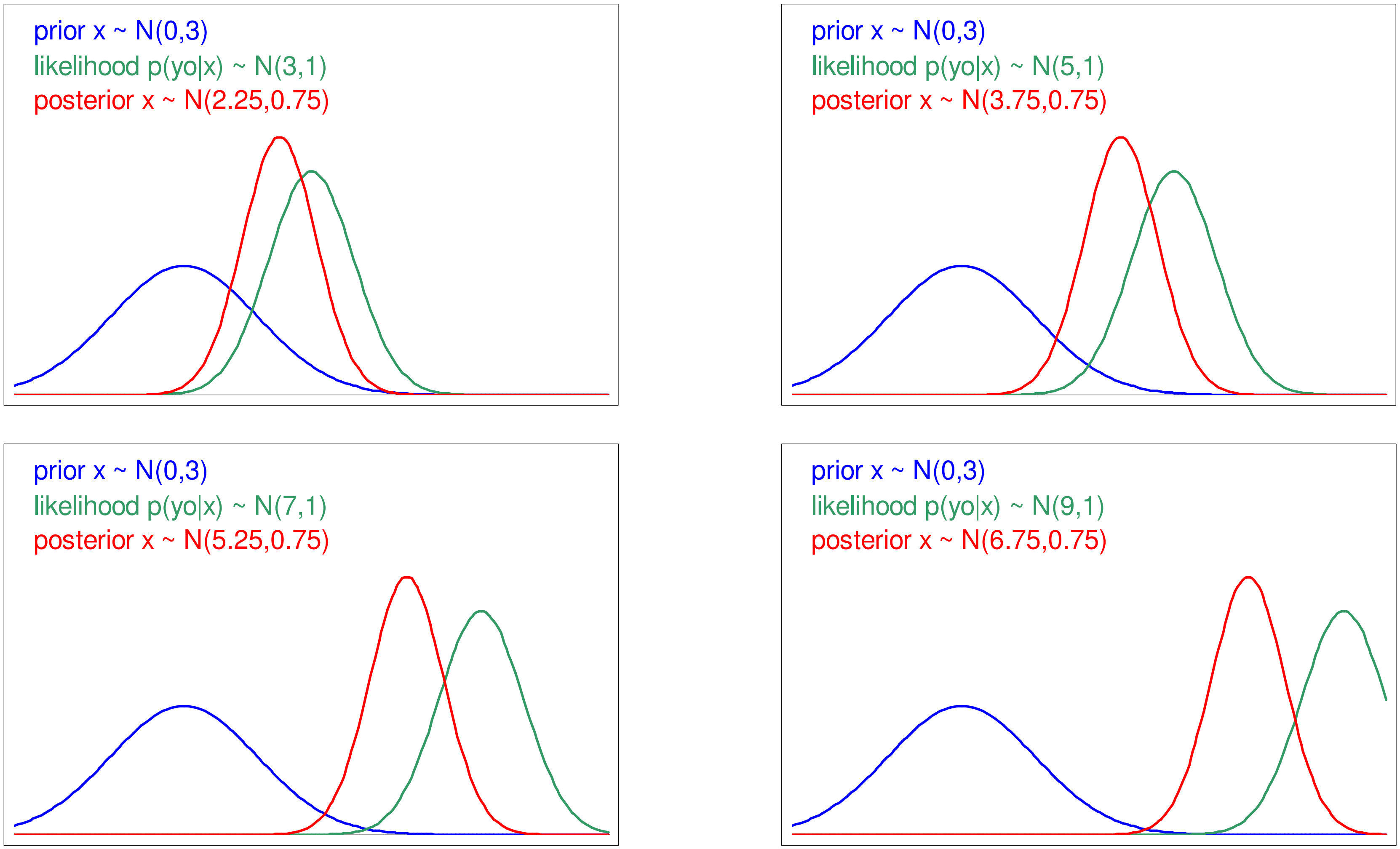
Taking logs gives quadratic equation; differentiating to find extremum gives linear equation.

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.



Quality Control example: Bayesian Dice

Discrete Bayes Theorem Applied to Gross Observational Errors

I have two dice. One is weighted towards throwing sixes. I have performed some experiments with them, and have the prior statistics that:

for the weighted (W) die, $P(6|W) = 58/60$

for the good (G) die, $P(6|G) = 10/60$

I choose one at random: $P(W) = P(G) = ½ = 50\%$

I throw this die, and it shows a six. Now:-

$$P(6) = P(6|W)\ P(W) + P(6|G)\ P(G$$
$$= 58/60 \quad 1/2 \quad + \quad 10/60 \quad 1/2$$
$$= 34/60$$

We can now apply Bayes' Theorem:

$$P(G|6) = P(6|G)\ P(G)\ /\ P(6)$$
$$= 10/60 \quad 1/2 \quad / \ 34/60 \ = \ 5/34 \ = 15\%$$
$$P(W|6) = P(6|W)\ P(W)\ /\ P(6)$$
$$= 58/60 \quad 1/2 \quad / \ 34/60 = 29/34 = 85\%$$

Simple model for PDF of observations with errors

Assume that a small fraction of the observations are corrupted, and hence worthless. The others have Gaussian errors.

For each observation we have:

$$p\left(y^o|x\right) = p\left(y^o|G \cap x\right)P(G) + p\left(y^o|\overline{G} \cap x\right)P\left(\overline{G}\right)$$

G is the event "there is a gross error" and $\overline{G}$ means *not* G.

$$p\left(y^o|\overline{G} \cap x\right) = N\left(y^o|H(x), \mathrm{E}+\mathrm{F}\right)$$

$$p\left(y^o|G \cap x\right) = \begin{cases} k & \text{over the range of plausible values} \\ 0 & \text{elsewhere} \end{cases}$$

Met Office

Applying this model

- Can simply apply Bayes Theorem to the discrete event G

$$P(G \mid y^o) = \frac{P(y^o \mid G)P(G)}{P(y^o)}$$

$$= \frac{k\, P(G)}{k\, P(G) + N\left(y^o \mid H(x^b), \mathbf{R} + H\mathbf{B}H^T\right) P(\overline{G})}$$

Lorenc, A.C. and Hammon, O., 1988: "Objective quality control of observations using Bayesian methods. Theory, and a practical implementation." *Quart. J. Roy. Met. Soc.*, **114**, 515-543

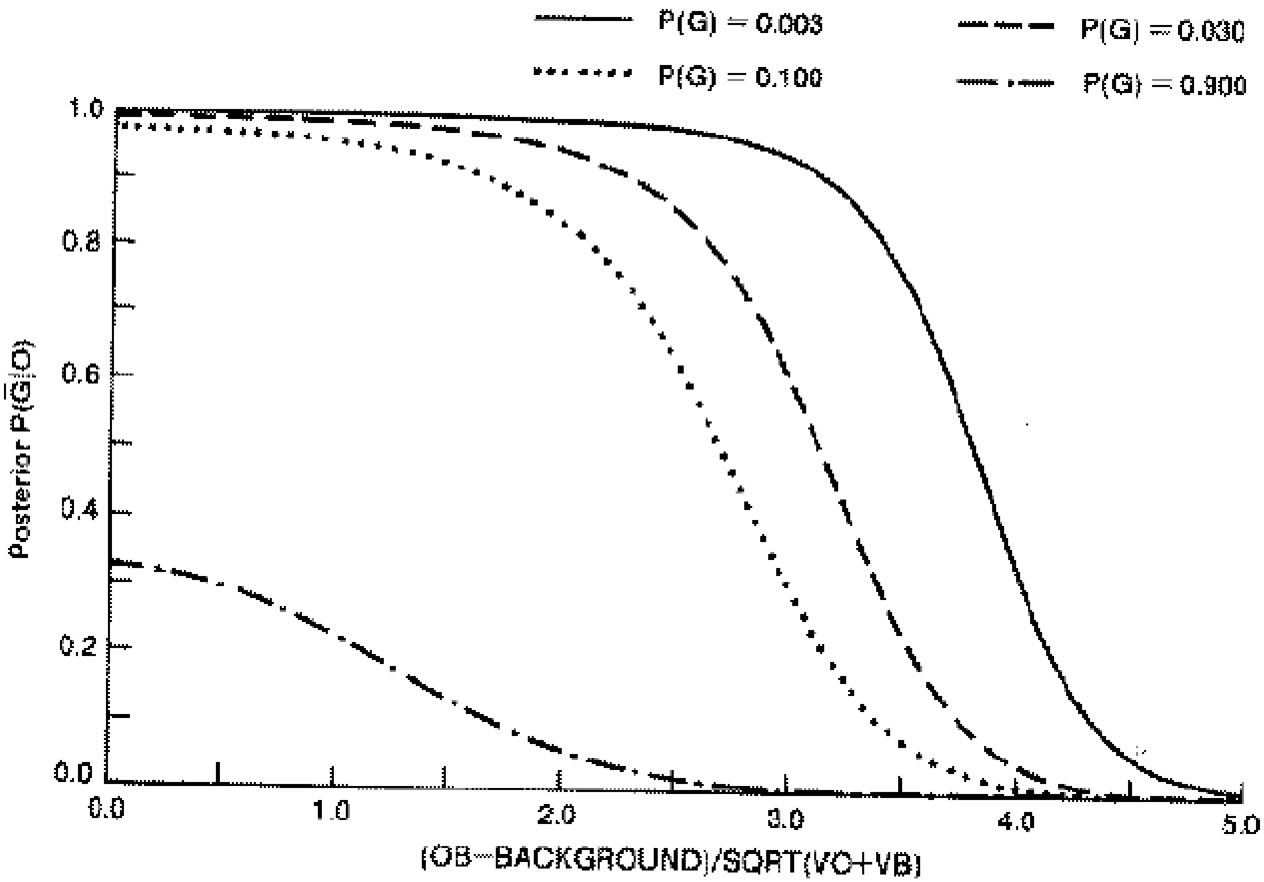
- Or we can use the non-Gaussian PDF directly

$$p(y^o \mid x) = p(y^o \mid G \cap x)P(G) + p(y^o \mid \overline{G} \cap x)P(\overline{G})$$

Ingleby, N.B., and Lorenc, A.C. 1993: "Bayesian quality control using multivariate normal distributions". *Quart. J. Roy. Met. Soc.,* **119**, 1195-1225

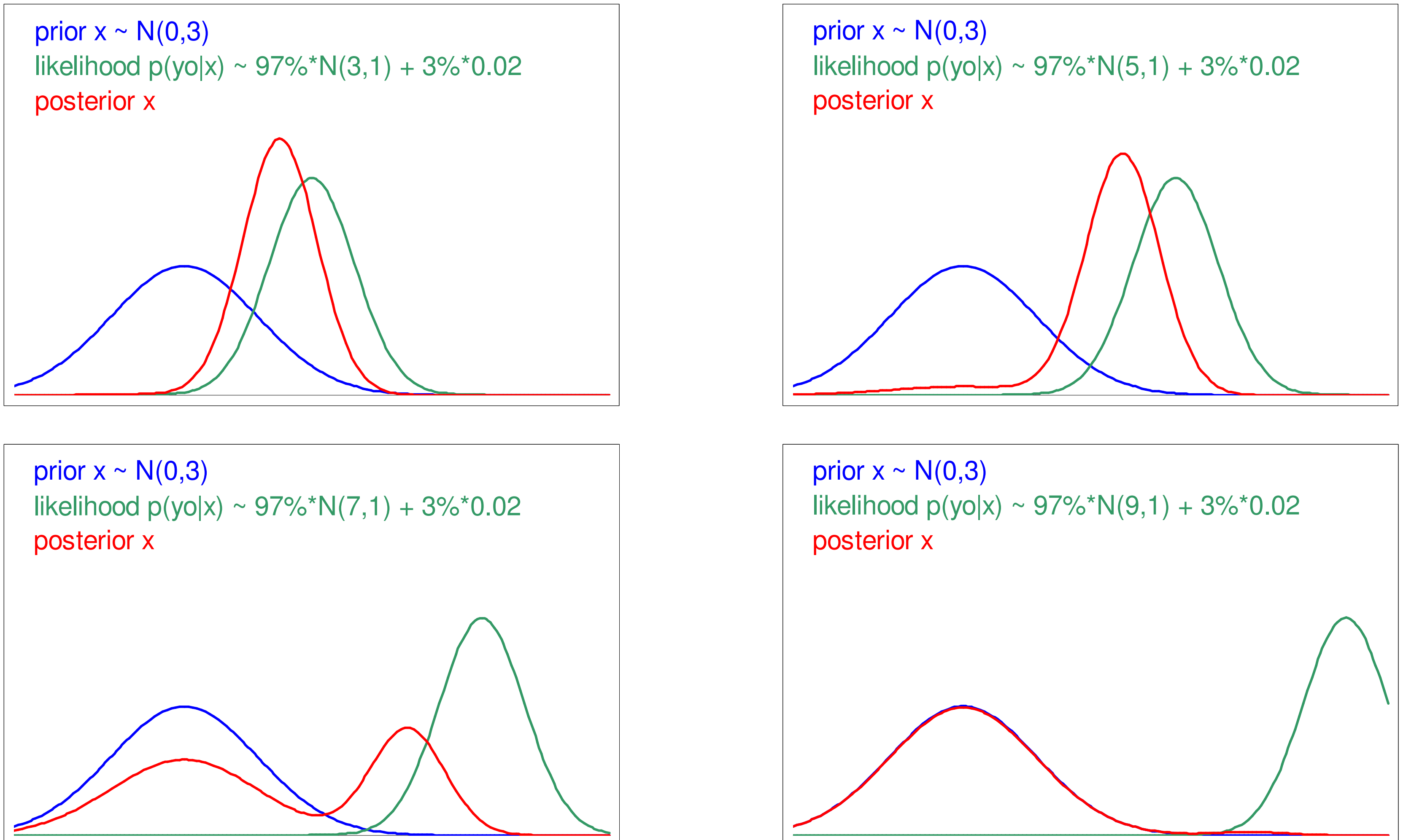
Andersson, Erik and Jarvinen, Heikki. 1999: "Variational Quality Control" *Quart. J. Roy. Met. Soc.*, **125**, 697-722

Posterior probability that an observation is "correct", as a function of its deviation from the background forecast



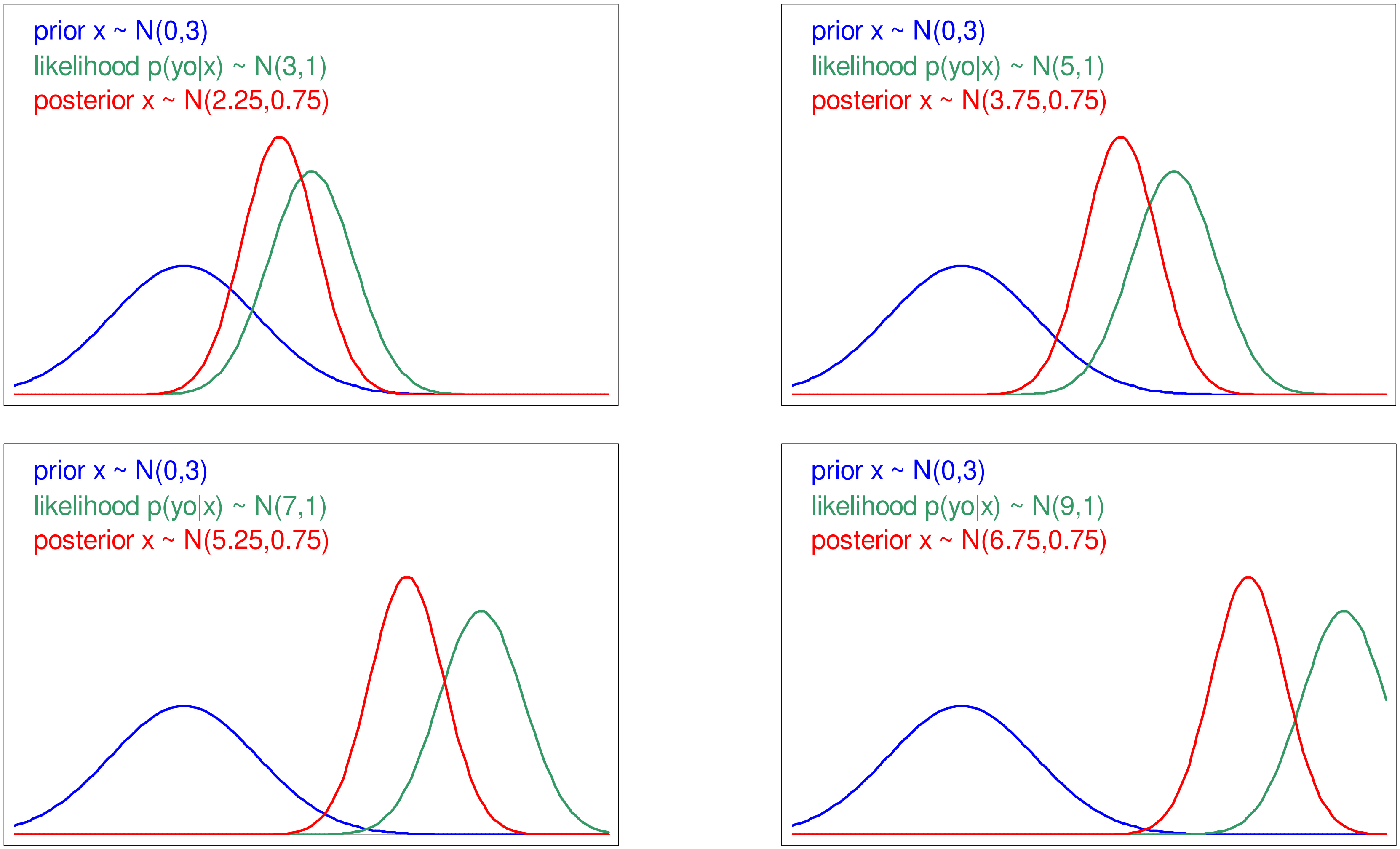
Posterior probability of an observation not having a gross error, plotted against normalized observed minus background value, for various prior probabilities of gross error.

Gaussian prior combined with observation with gross errors - extreme obs are rejected.



Js

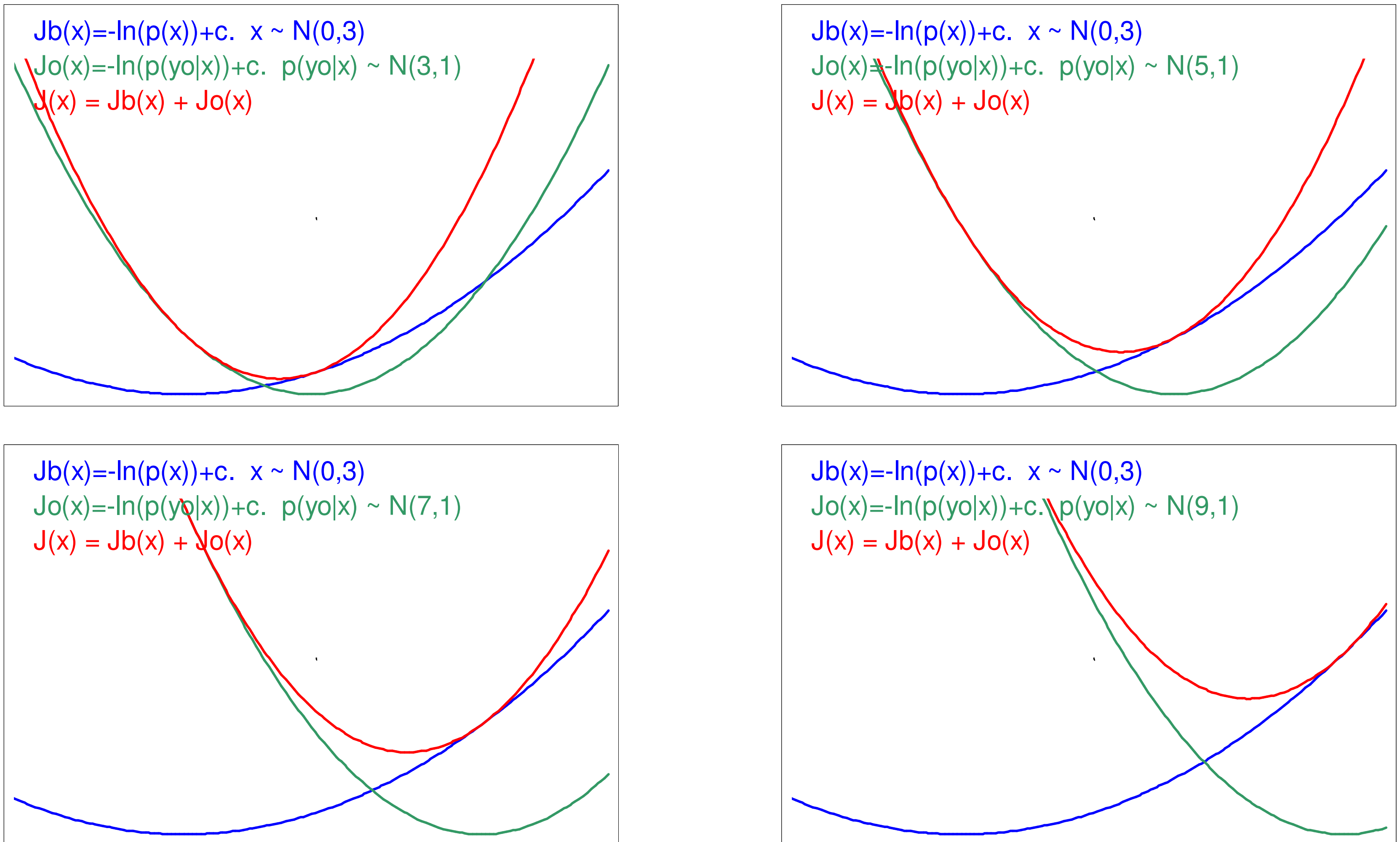
Combination of Gaussian prior & observation
- Gaussian posterior,
- weights independent of values.



Variational Penalty Functions

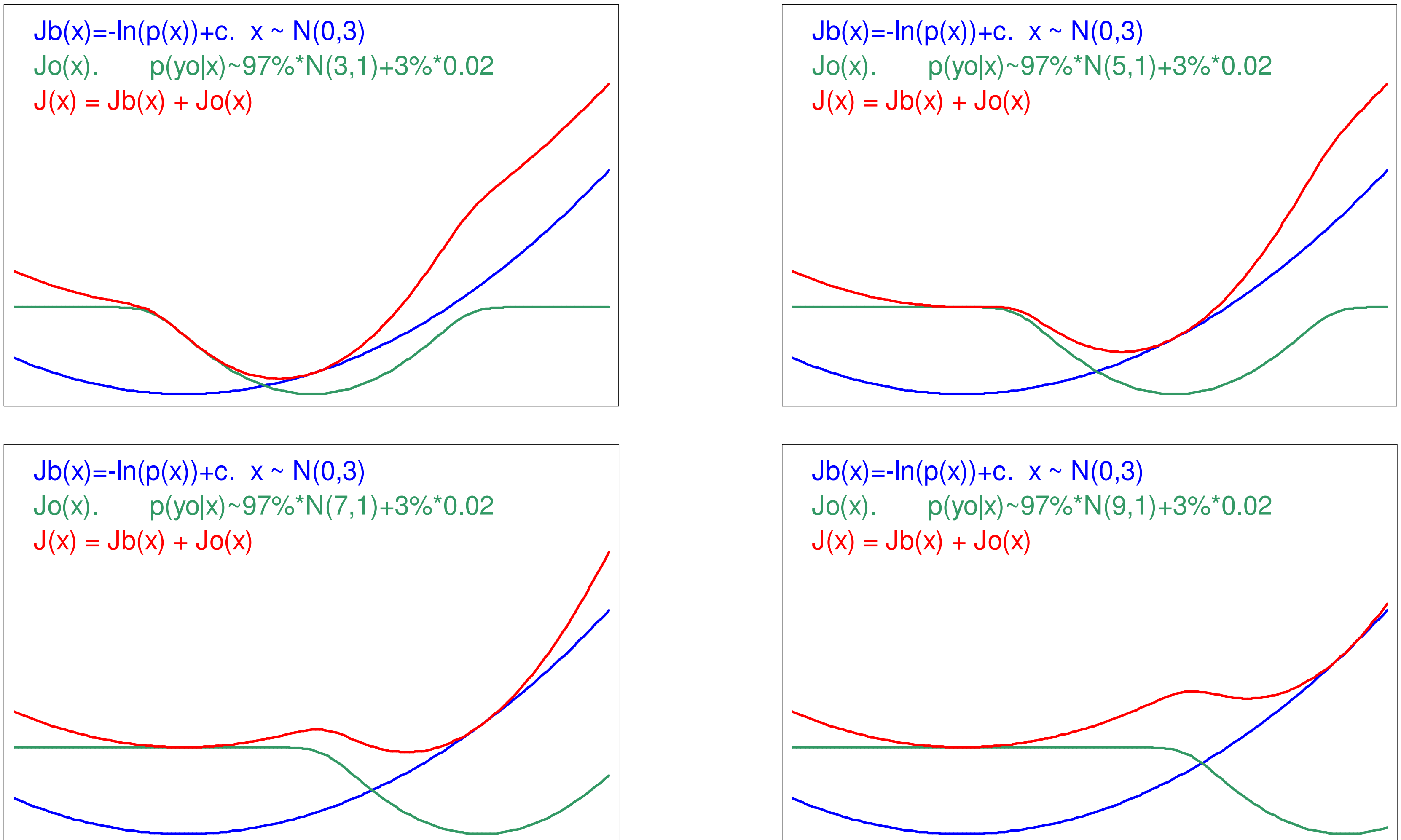
- Finding the most probable posterior value involves maximising a product [of Gaussians]
- By taking $-ln$ of the posterior PDF, we can instead minimise a sum [of quadratics]
- This is often called the “Penalty Function” J
- Additive constants can be ignored

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

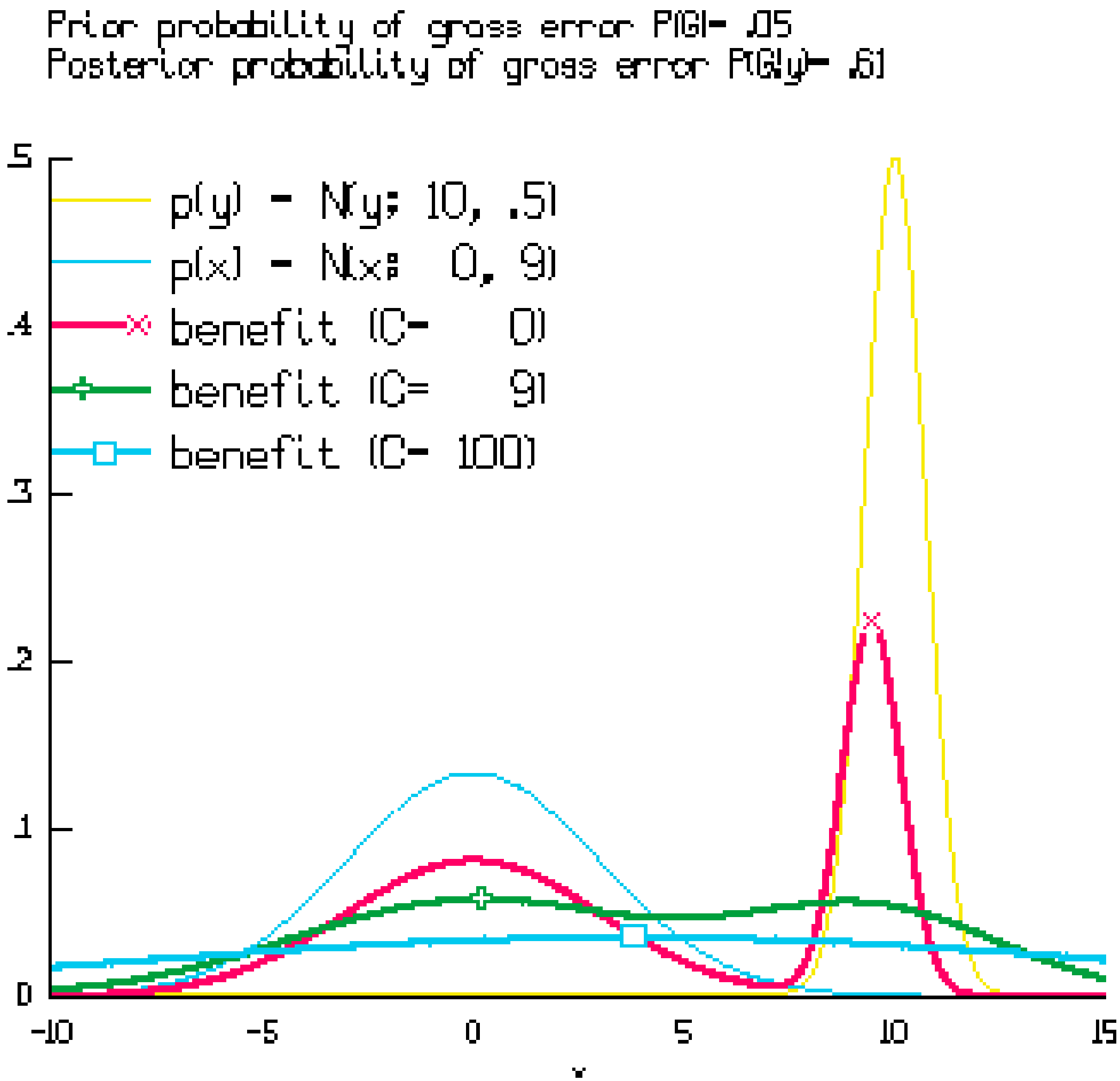


PDFs

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



PDFs



Expected benefit as a function of analysed value. Curves are plotted for three different benefit functions, with widths C=0 (maximum at **X**), C=9 (maximum at **+**), C=100 (max at ☐). Shown for reference are the background pdf (with x^b=0, B=9), and the observational pdf (with y^o=10, R=0.5).

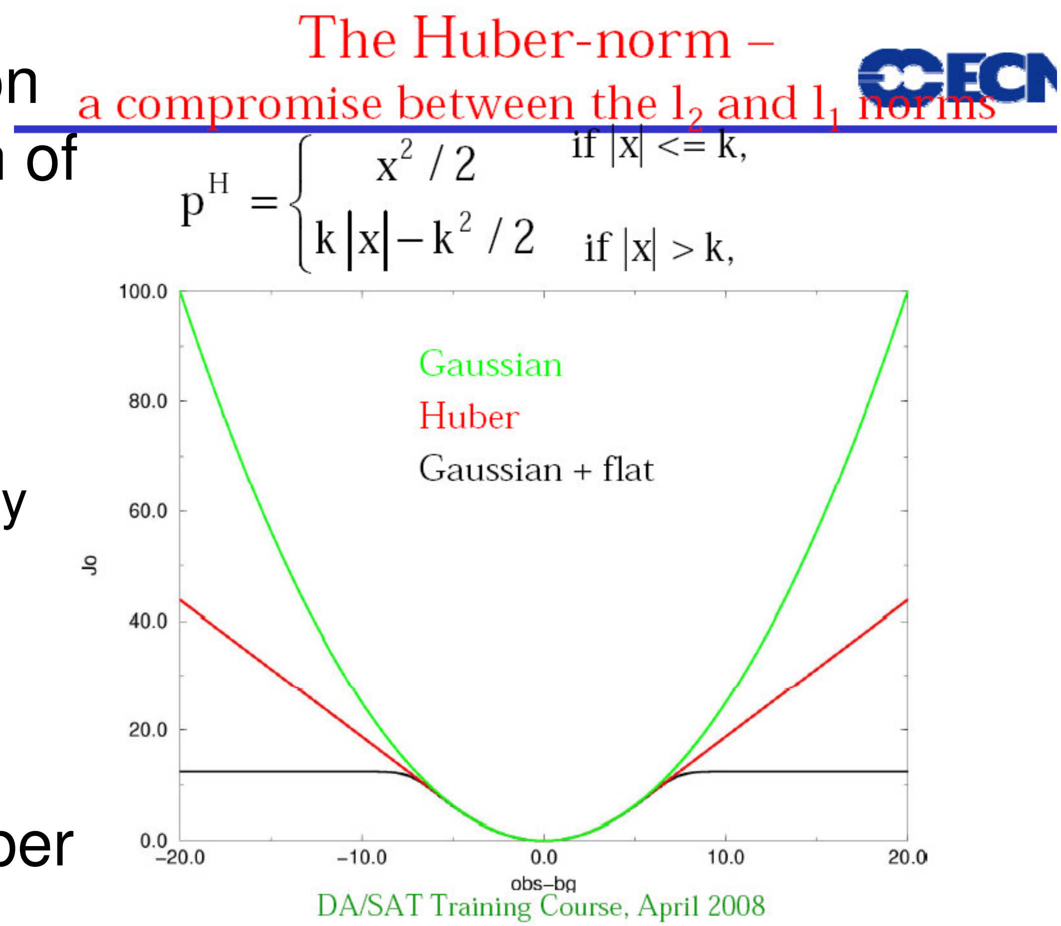
Lorenc, A. C., 2002: Atmospheric Data Assimilation and Quality Control. *Ocean Forecasting*, eds Pinardi & Woods. ISBN 3-540-67964-2. 73-96

Other models for observation error

The Gaussian + flat distribution lead to a rather rapid rejection of observations with large deviations. This can cause problems:

1. When the guess is some way from the observation
2. When the observation is of (important) severe weather

Erik Andersson prefers a "Huber Norm" penalty function:



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): $\mathbf{y}^o = \left(y^o \right)$

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

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

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:

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

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

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

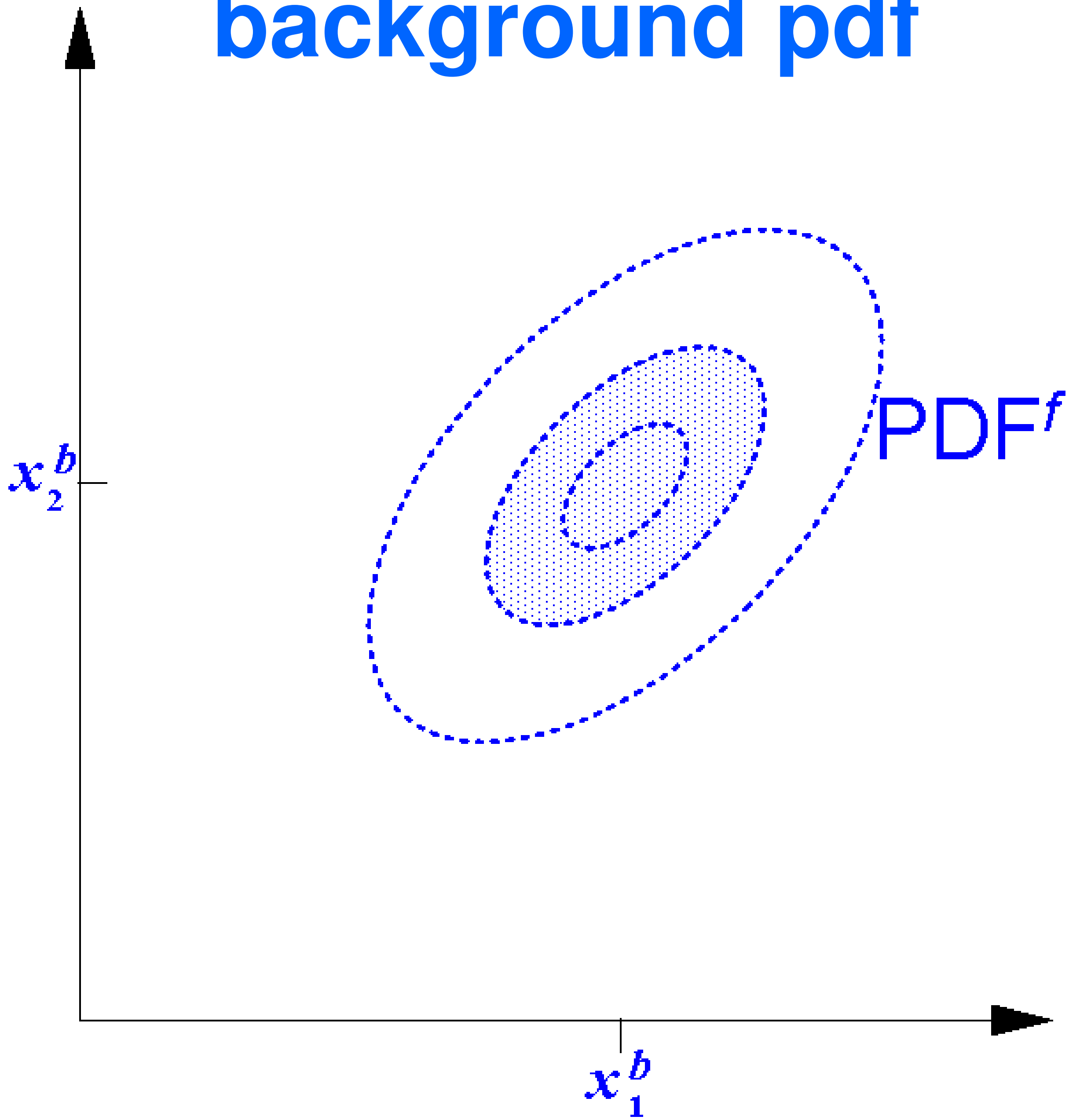
$$\mathbf{x} \sim N(\mathbf{x} : \mathbf{x}^b, \mathbf{B})$$

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

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

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

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

error of representativeness

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

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

Observational error combines these 2 :

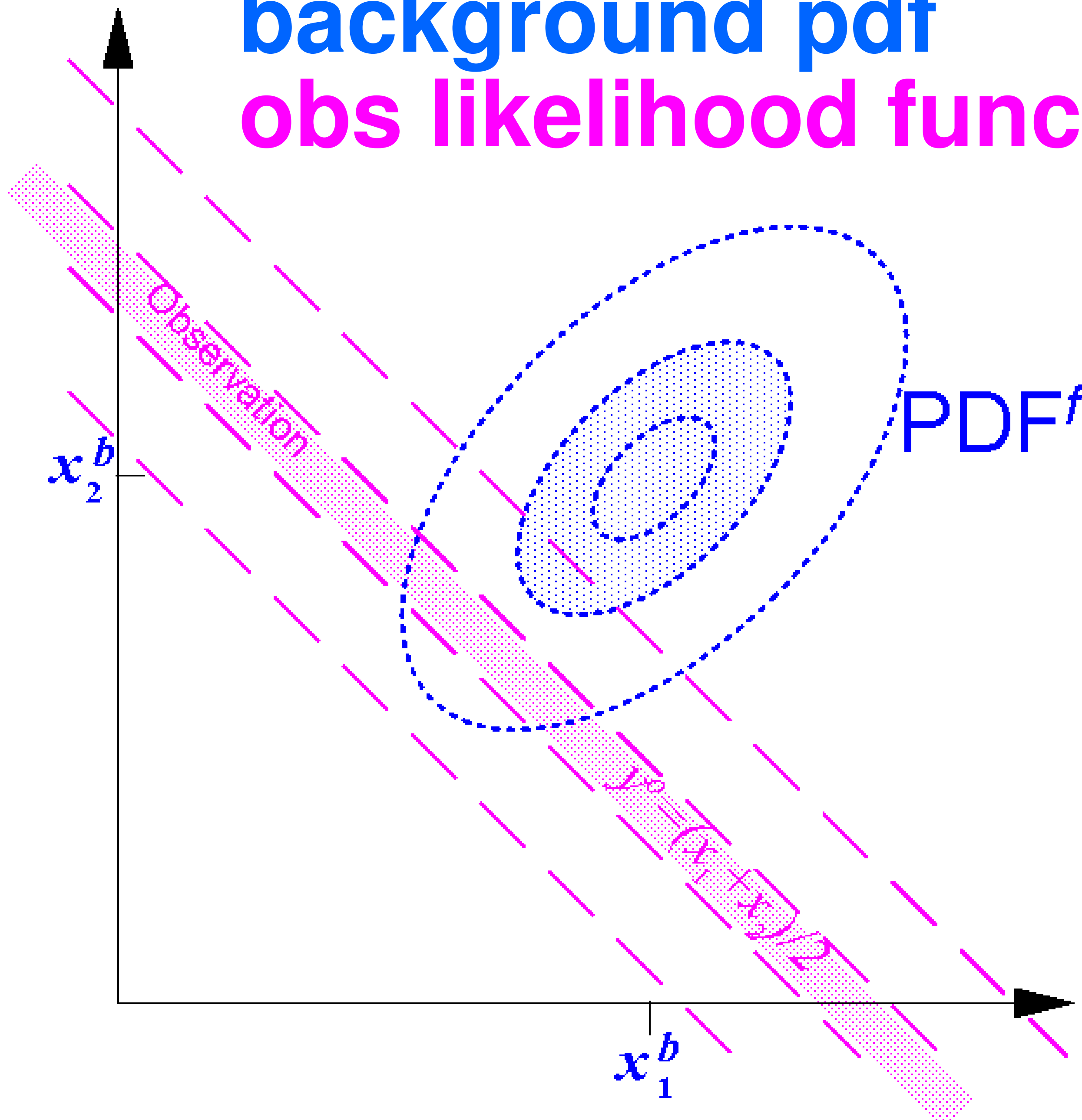
$$\mathbf{y}^o \sim N\left(H\left(\mathbf{x}^t\right), \mathbf{E}+\mathbf{F}\right)$$

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

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

Met Office

background pdf
obs likelihood function



Bayesian analysis equation

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

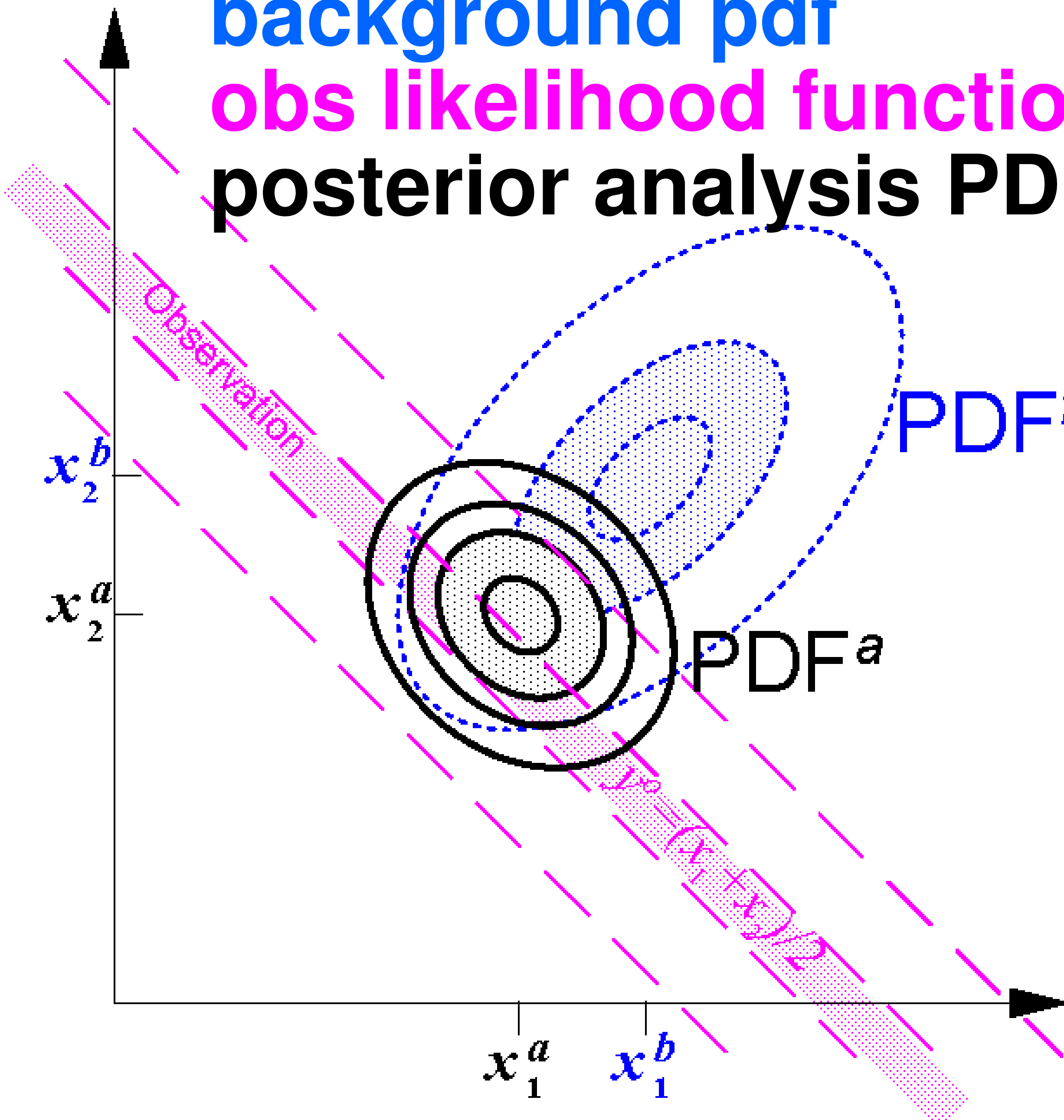
Property of Gaussians that, if H is linearisable : $\mathbf{x} \sim N(\mathbf{x}^a, \mathbf{A})$

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

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

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

background pdf
obs likelihood function
posterior analysis PDF



Analysis equation

For our simple example the algebra is easily done by hand, giving:

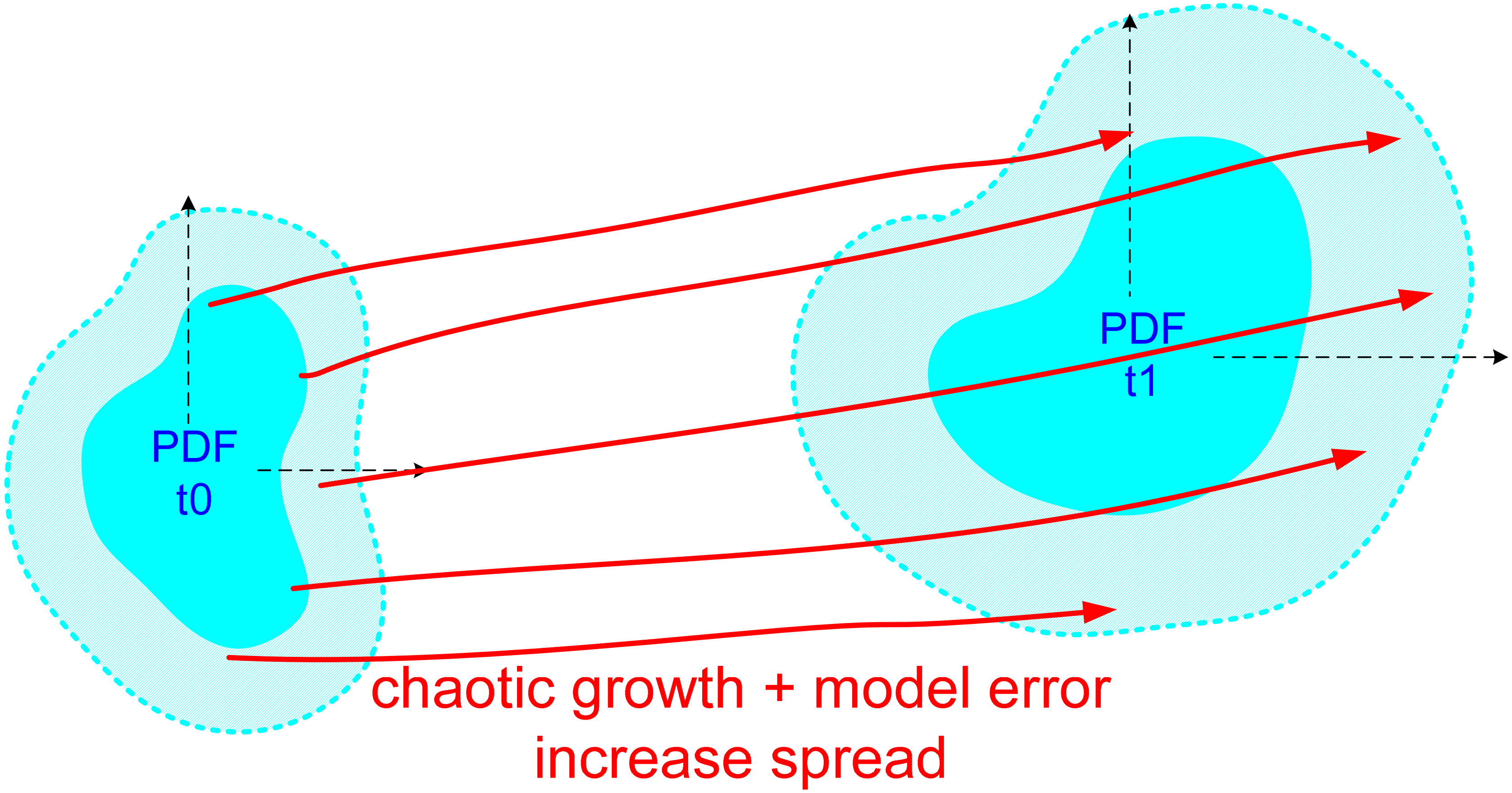
$$\mathbf{x}^a = \begin{pmatrix} x_1^a \\ \\ x_2^a \end{pmatrix} = \begin{pmatrix} x_1^b \\ \\ x_2^b \end{pmatrix} + \frac{\left(V^b\left(\frac{1+\mu}{2}\right)\right)^2}{\mathbf{E}+\mathbf{F}+V^b\left(\frac{1+\mu}{2}\right)}\left[y^o - \frac{x_1^b + x_2^b}{2}\right]\begin{pmatrix} 1 \\ \\ 1 \end{pmatrix}$$

Met Office

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

i.e. 4D-Var versus Ensemble KF

Fokker-Planck Equation

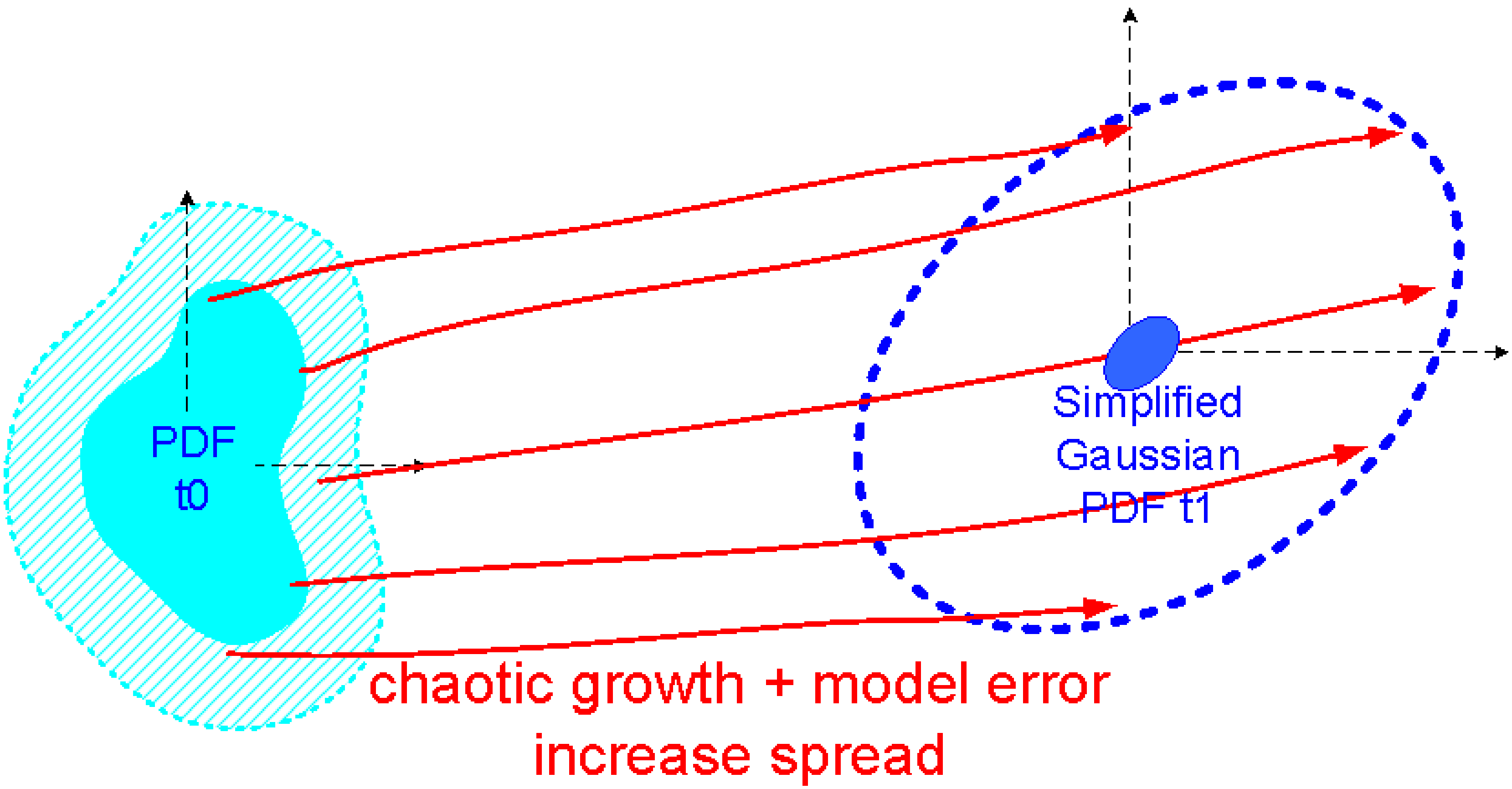


Ensemble methods attempt to sample entire PDF.

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)

Ensemble Kalman filter



Fit Gaussian to forecast ensemble.

The Ensemble Kalman Filter (EnKF)

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

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

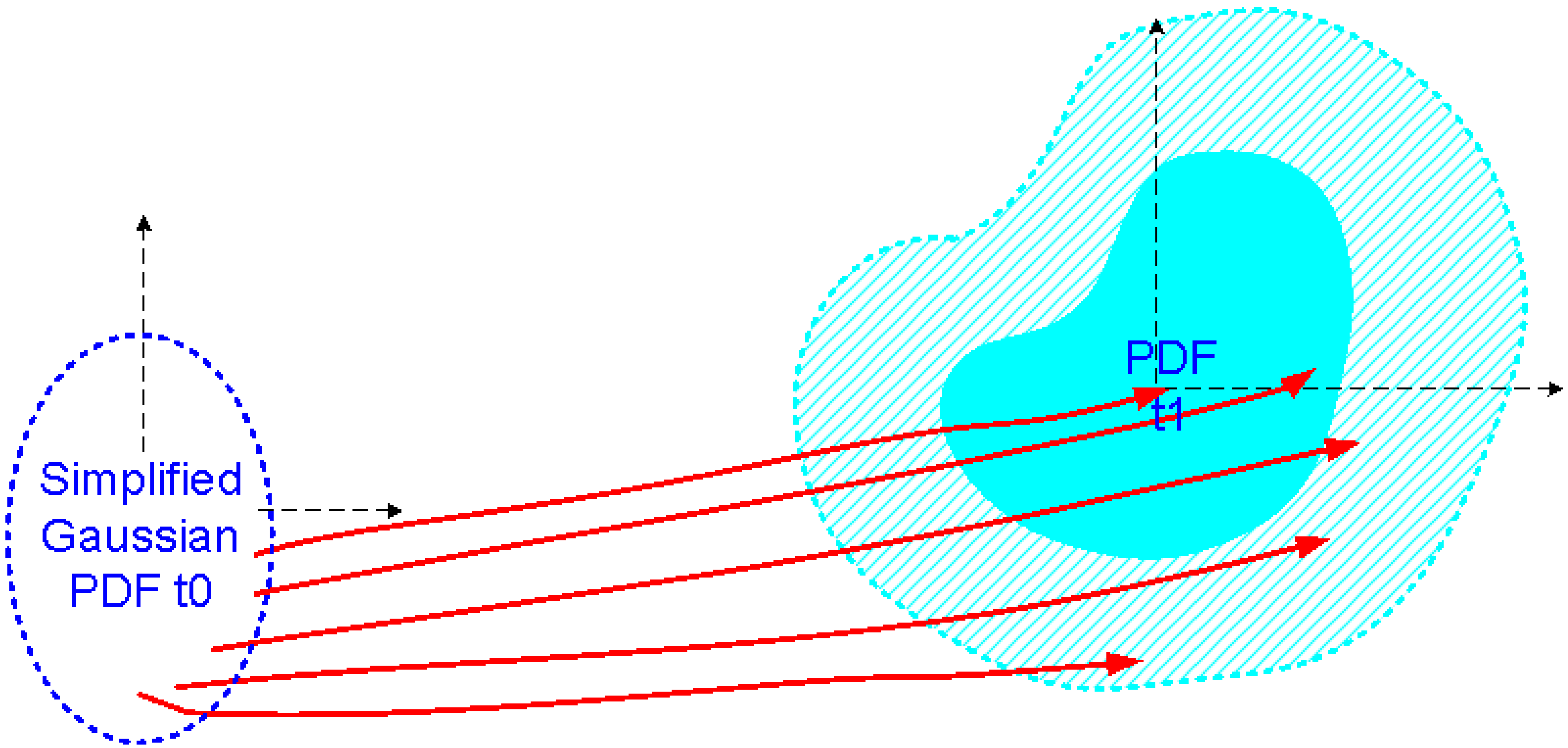
$$\mathbf{P}^f\mathbf{H}^T = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)^T},$$

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

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

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

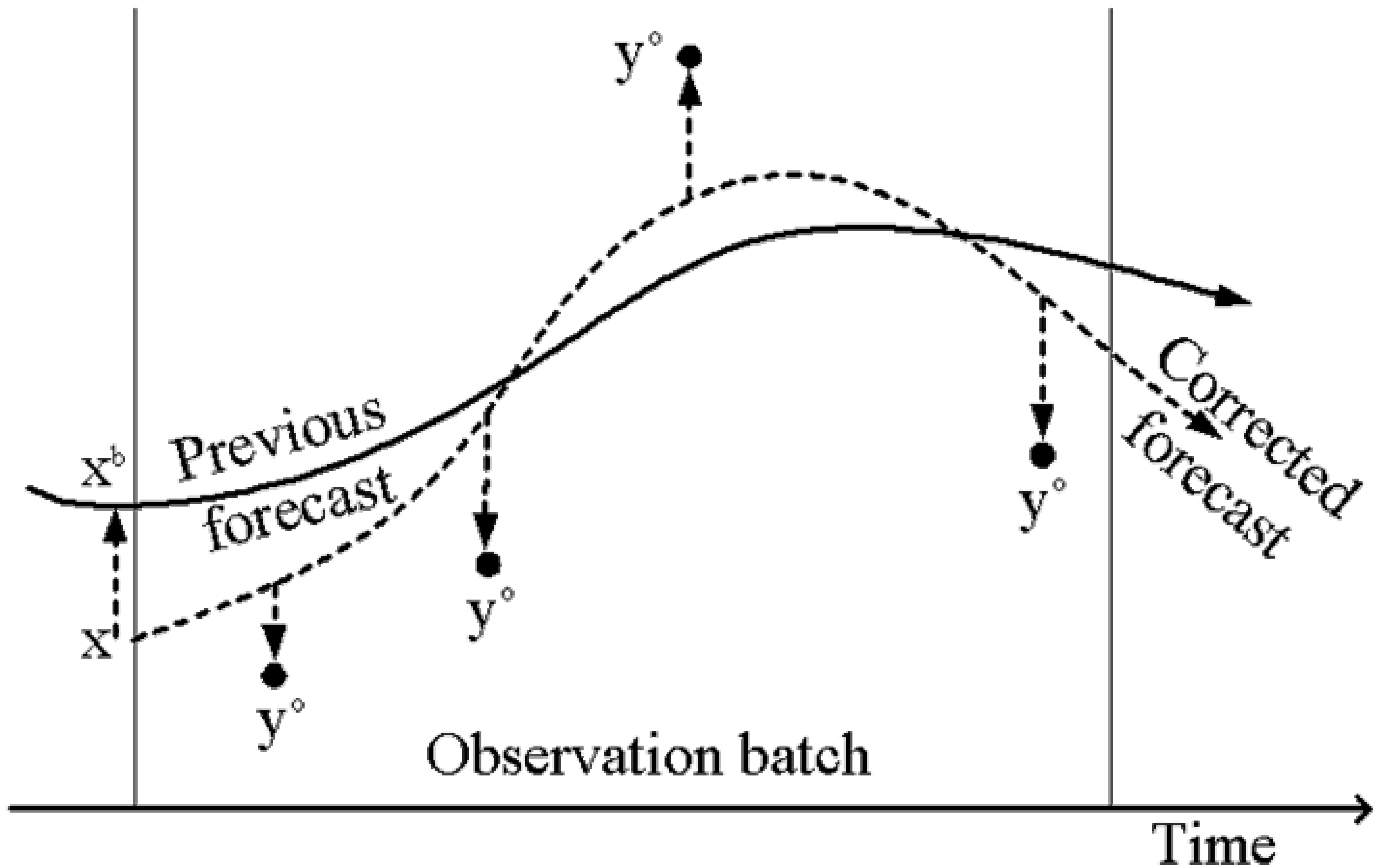
Deterministic 4D-Var



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

Met Office

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

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

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

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

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.

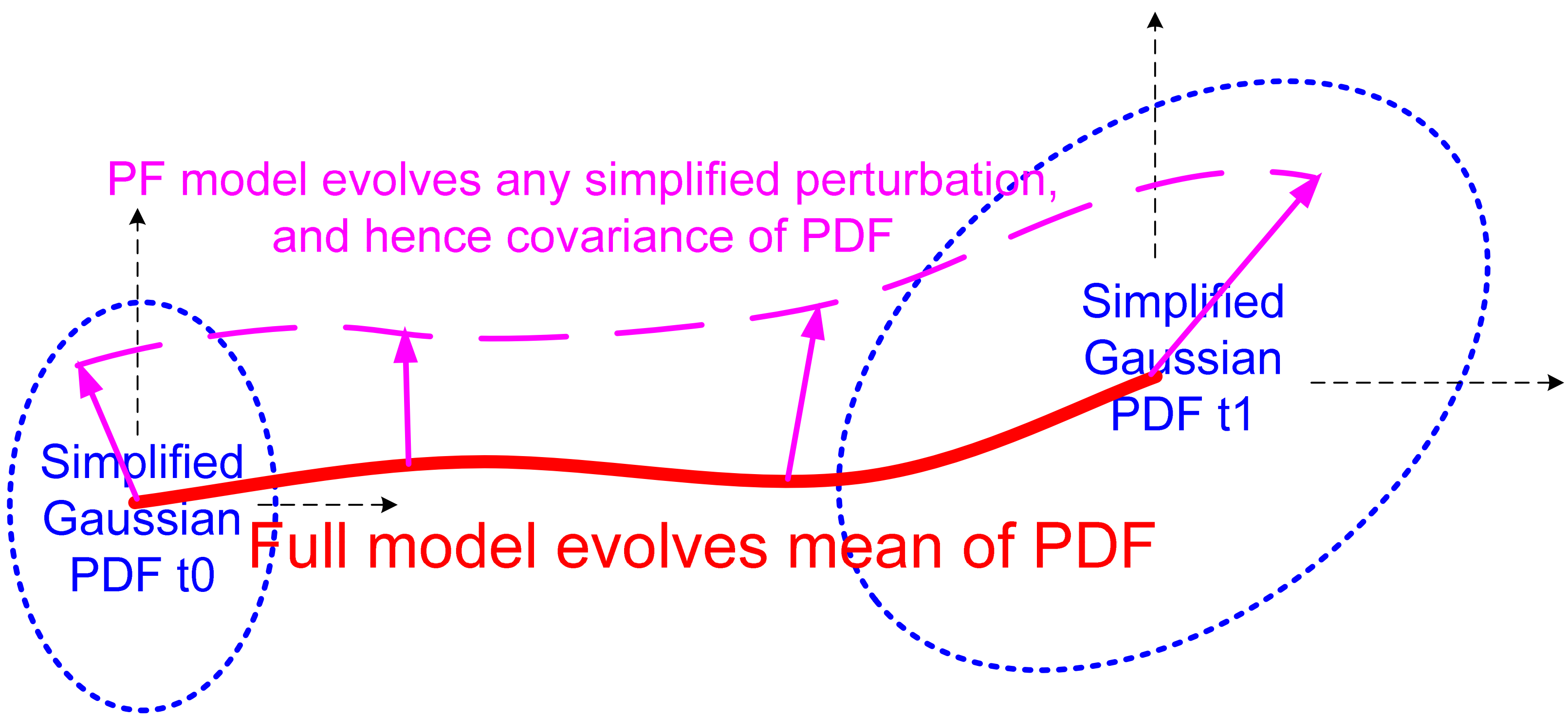
$$\underline{\mathbf{y}} = \underline{H}\left(\underline{M}(\mathbf{x})\right)$$

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:

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

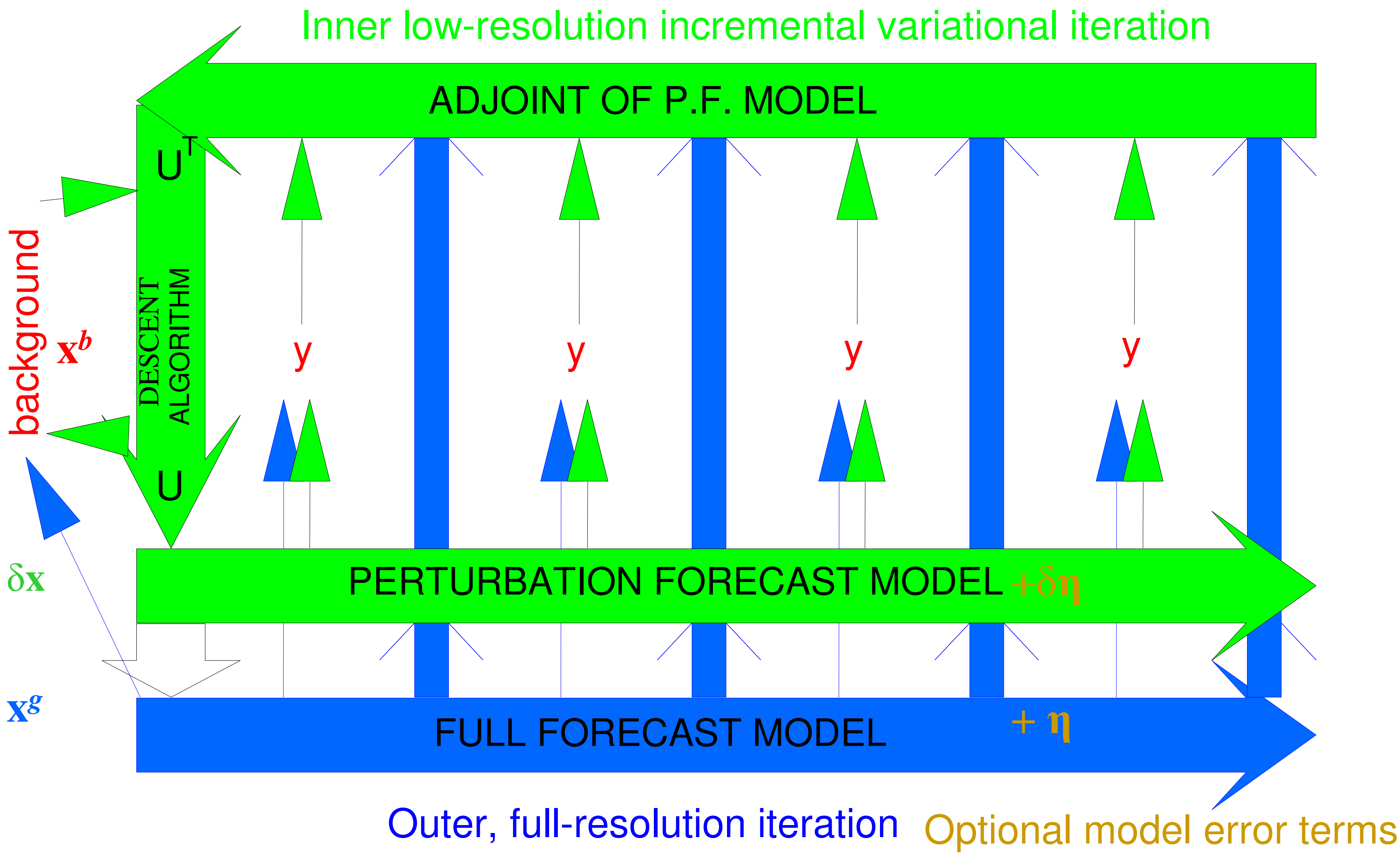
The mean of this approximate pdf is identical to the mode, so it can be found by minimising:

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

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

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

Incremental 4D-Var with Outer Loop



Questions and answers

Content

1. Bayes Theorem – adding information
    - Gaussian PDFs
    - Non-Gaussian observational errors - Quality Control
2. Simplest possible Bayesian NWP analysis
    - Two gridpoints, one observation.
3. Predicting the prior PDF

    – a Bayesian view of 4D-Var v Ensemble KF